

Friedrich-Alexander-Universität Erlangen-Nürnberg

**Lehrstuhl für Multimediakommunikation und
Signalverarbeitung**

Prof. Dr.-Ing. André Kaup

Projektarbeit

**Analyse von Qualitätsmetriken für die
Evaluierung der Qualität des Tiefpassbandes
einer kompensierten Wavelet-Transformation**

von B. Sc. Susanne Schöll

14.10.2013

Betreuer: Dipl.-Ing. Wolfgang Schnurrer

Projektarbeit

für

Frau Susanne Schöll

Analyse von Qualitätsmetriken für die Evaluierung der Qualität des Tiefpassbandes einer kompensierten Wavelet-Transformation


Das Tiefpassband der Wavelet-Transformation kann als herunterskalierte Version des originalen Signals verwendet werden. Dafür ist eine hohe Qualität des Tiefpassbandes sehr wichtig. Bei einer Wavelet-Transformation in zeitlicher Richtung können Bewegungen im Signal jedoch zu einem unscharfen Tiefpassband mit Bewegungsartefakten und Geisterbildern führen. Diese Artefakte können durch die Erweiterung der Transformation um Kompensationsverfahren reduziert werden. In der Videocodierung kommen üblicherweise blockbasierte Verfahren zum Einsatz. Während der Wavelet-Transformation muss das Kompensationsverfahren invertiert werden. Bei blockbasierten Verfahren können dabei Löcher entstehen, die im Tiefpassband zu störenden Artefakten führen können.

Die gängige Beurteilung der Qualität vergleicht das Tiefpassband mit dem sogenannten *korrespondierenden* originalen Signalanteil. Als Metrik wird üblicherweise das Spitzen-Signal-zu-Rauschleistungsverhältnis (PSNR) gemessen. Diese Messung berücksichtigt allerdings die zuletzt beschriebenen Artefakte nicht. Eine optimale Metrik sollte dagegen zum einen die Skalierbarkeit und zum anderen die visuelle Qualität berücksichtigen. Da das Tiefpassband das originale Signal möglichst vollständig repräsentieren soll, ist im Allgemeinen nicht klar, welcher Anteil des originalen Signals repräsentativ als Referenz für die Qualitätsbeurteilung verwendet werden kann. In dieser Arbeit sollen daher eine Metrik untersucht werden, die keine Referenz für die Qualitätsbeurteilung verwendet.

Frau Schöll erhält die Aufgabe die referenzfreie Qualitätsmetrik Naturalness Image Quality Evaluator (NIQE) aus der Literatur hinsichtlich ihrer Eignung zur Bewertung der Qualität des Tiefpassbandes zu untersuchen. Zur Bewertung soll die Metrik für die Tiefpassbänder verglichen werden, die sich aus der Wavelet-Transformation ohne Kompensation, mit blockbasierter Kompensation sowie mit blockbasierter Kompensation und zusätzlicher Lochfüllung resultieren. Außerdem soll untersucht werden, in wie weit die Metrik die Skalierbarkeit bezüglich der Rate abbildet. Zur Auswertung sollen Videosequenzen sowie medizinische Volumendaten verwendet werden.

Im Rahmen der Arbeit soll eine Literaturrecherche zur Einordnung der zu untersuchenden Metrik durchgeführt werden. Auf saubere Dokumentation der Arbeit und des Quelltextes wird besonderer Wert gelegt.

Beginn: 15.04.2013


(Prof. Dr.-Ing. A. Kaup)

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe, und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Ort, Datum

Unterschrift

Inhaltsverzeichnis

Kurzfassung	IV
Formelzeichen	VII
1 Einleitung	1
2 Grundlagen Qualitätsbeurteilung	3
2.1 Einführung zum Thema	3
2.2 Berechnung des NIQE-Algorithmus	4
2.2.1 Spatial Domain NSS	5
2.2.2 Patch Selection	5
2.2.3 Characterizing Image Patches	6
2.2.4 Multivariate Gaussian Model	6
2.2.5 NIQE-Index	7
2.3 Modell Generierung für NIQE	7
2.3.1 Modell aus Datenbank	7
2.3.2 Modell aus CT-Datenbank	8
2.3.3 Modell aus aktueller Sequenz bzw. bestimmter Schicht von CT-Daten	9
3 Berechnung der Tiefpassbänder aus der Wavelet-Transformation	10
3.1 Tiefpassband ohne Kompensation	12
3.2 Tiefpassband mit blockbasierter Kompensation	13
3.3 Tiefpassband mit blockbasierter Kompensation und Lochfüllung	14

4	Auswertung der Ergebnisse	16
4.1	Simulationssetup / Simulationsdaten	16
4.1.1	Basketball - Beschreibung	16
4.1.2	Discovery-City - Beschreibung	18
4.1.3	Crew - Beschreibung	18
4.1.4	Flower - Beschreibung	19
4.1.5	Foreman - Beschreibung	19
4.1.6	Discovery-Orient - Beschreibung	19
4.2	NIQE mit Datenbank-Modell im Vergleich mit PSNR	21
4.2.1	Basketball	21
4.2.2	Discovery-City	27
4.2.3	Crew	34
4.2.4	Flower	44
4.2.5	Foreman	49
4.2.6	Discovery-Orient	55
4.2.7	Head	61
4.3	NIQE mit CT-Datenbank-Modell	65
4.3.1	Head-Trainierter NIQUE vs. NIQUE aus der Veröffentlichung	66
4.3.2	Thorax-Trainierter NIQUE vs. NIQUE aus der Veröffentlichung	69
4.3.3	Neck-Trainierter NIQUE vs. NIQUE aus der Veröffentlichung	72
4.4	NIQE mit speziell trainiertem Modell	74
4.4.1	Videosequenz Crew	75
4.4.2	Schicht 10 der CT-Volumendaten Neck	78
5	Zusammenfassung der Ergebnisse zur neuen Metrik	81
	Abbildungsverzeichnis	84
	Tabellenverzeichnis	86
	Literaturverzeichnis	88

Kurzfassung

Zur Reduzierung von irrelevanten und redundanten Bilddaten in der digitalen Signalverarbeitung kann die Wavelet-Transformation (WT) herangezogen werden. Das hieraus resultierende Tiefpassband (TP) einer WT kann als Vorschau mit niedrigerer Auflösung verwendet werden. Für den weiteren Gebrauch ist allerdings eine hohe Qualität des TPs unabdingbar. Um das zu gewährleisten, wird die WT um verschiedene Kompensationsverfahren erweitert, wodurch entstehende Ghosting-Artefakte reduziert werden können.

In dieser Projektarbeit werden verschiedene TPs zur Qualitätsbeurteilung eingesetzt. Hierfür werden zum einen das TP aus der WT ohne Kompensation (L-zero) sowie das TP mit blockbasierter Kompensation (L-block) berechnet. Des Weiteren soll ein TP mit blockbasierter Kompensation und zusätzlicher Lochfüllung (L-FSE) verwendet werden, wozu die Methode der Frequency Selective Extrapolation (FSE) berücksichtigt wird.

Die gewöhnliche Qualitätsbeurteilung erfolgt anhand einer Gegenüberstellung des TPs und dem zugehörigen Originalsignal, wie bei der bisher üblichen Metrik, dem Peak-Signal-to-Noise-Ratio (PSNR), verfahren wird. Da sich diese Qualitätsbewertung stets auf eine Referenz bezieht, handelt es sich bei dem PSNR um eine Full-Reference (FR) Metrik. In dieser Arbeit soll ein anderer Ansatz untersucht werden, bei dem der Natural Image Quality Evaluator (NIQE) bewertet wird. Das Besondere an diesem Algorithmus ist, dass er keine Referenz benötigt, um die Qualität eines Bildes zu bestimmen. Der Aspekt der referenzfreien Qualitätsbewertung zeichnet den NIQE als No-Reference (NR) Metrik aus.

Um die Eignung des NIQEs hinsichtlich der Qualitätsbewertung zu überprüfen, werden unterschiedliche Videosequenzen sowie medizinische Volumendaten zur Auswertung herangezogen und die Ergebnisse mit dem PSNR verglichen.

Abkürzungsverzeichnis

Abb.	Abbildung
CT-Daten	Computertomografie-Daten
DU	Distortion Unaware
FR	Full-Reference
FSE	Frequency Selective Extrapolation
IQA	Image Quality Assessment
L-block	Tiefpassband mit blockbasierter Kompensation
L-FSE	Tiefpassband mit blockbasierter Kompensation und Lochfüllung
L-zero	Tiefpassband ohne Kompensation
LMS	Lehrstuhl für Multimediakommunikation und Signalverarbeitung
MVG-Modell	Multivariates Gauß Modell
NIQE	Natural Image Quality Evaluator
NR	No-Reference
NSS	Natural Scene Statistic
OU	Opinion Unaware

PSNR	Peak-Signal-to-Noise-Ratio
R-orig	Referenzbild Original
Tab.	Tabelle
TP	Tiefpassband
WT	Wavelet-Transformation

Formelzeichen

- | | | |
|------|----------------------------------|---|
| (1) | $PSNR_V$ | PSNR für Videosequenzen |
| (2) | $PSNR_{CT}$ | PSNR für CT-Volumen |
| (3) | \hat{I} | Koeffizient |
| (4) | μ | Mittelwert |
| (5) | σ | Varianz |
| (6) | δ | lokale Schärfe |
| (7) | $f(x; \alpha, \beta)$ | allgemeine Gaußverteilung |
| (8) | Γ | Gamma Funktion |
| (9) | $f(x; \gamma, \beta_l, \beta_r)$ | asymmetrische standardisierte Gaußverteilung |
| (10) | f_x | MVG-Modell |
| (11) | D | NIQE-Index |
| (12) | HP_t | Hochpassband zum Zeitpunkt t |
| (13) | f_{2t} | aktuelles Bild zum Zeitpunkt $2t$ |
| (14) | f_{2t-1} | Referenzbild zum Zeitpunkt $2t - 1$ |
| (15) | LP_t | Tiefpassband zum Zeitpunkt t |
| (16) | p_{2t} | Prädiktor zum Zeitpunkt $2t$ |
| (17) | a_k | Gewichtungsfaktor |
| (18) | u_{2t} | traditionelles Update Frame zum Zeitpunkt $2t$ |
| (19) | \hat{u}_{2t} | rekonstruiertes Update Frame zum Zeitpunkt $2t$ |

Kapitel 1

Einleitung

Eine Methode, die sich in der Praxis durchgesetzt hat unwichtige Bildinformationen zu eliminieren, um mit einer geringeren Datenmenge zu arbeiten, stellt die Wavelet-Transformation (WT) dar. Dabei wird das Originalsignal in ein Hochpassband und in ein Tiefpassband (TP) geteilt. Das TP der WT kann aufgrund seiner geringeren Auflösung als eine herunterskalierte Originalversion bestimmt werden. JPEG2000 ist beispielsweise ein Standardverfahren zur Datenkomprimierung das die WT verwendet.

Bewegungen im Signal können aber zu einem unscharfen TP führen. Zur Verbesserung der Bildqualität von TPs kann die WT um verschiedene Kompensationsverfahren erweitert werden. In dieser Arbeit werden zur näheren Betrachtung zwei Arten von Kompensationsverfahren herangezogen. Zum einen handelt es sich um das TP aus der WT mit blockbasierter Kompensation, zum anderen um das TP aus der WT mit blockbasierter Kompensation und Lochfüllung sowie das TP aus der WT ohne Kompensation. Um letztendlich den Qualitätsunterschied zu verdeutlichen werden die Ergebnisse der TPs untereinander und mit dem, pro TP zugehörigen, originalen Referenzbild (R-orig) verglichen.

Die WT kann als Lifting Struktur effizienter berechnet werden und gliedert sich in zwei Schritte. Die Berechnung des Hochpassbandes erfolgt im Prediction Step, während im darauffolgenden Schritt, dem Update Step, das TP resultiert. Das Kompensationsverfahren wird im Update Step der WT invertiert. Bei der in der Videocodierung üblichen blockbasierten Kompensation kann dies zur Folge haben, dass ein TP mit Artefakten resultiert.

Für die Qualitätsbeurteilung von Bild- und Videosignalen wurde bislang das PSNR herangezogen. Es hat sich als Standardmetrik etabliert und liefert verlässliche Ergebnisse. Allerdings besteht beim PSNR eine für diese Arbeit entscheidende Problematik, was speziell die Bewertung der bei der WT resultierenden TPs betrifft. Diese Invertierung bei der blockbasierten Kompensation im Update Step führt zu unverbundenen Pixeln, auch Löcher genannt. An diesen Stellen fehlt sozusagen das Update zur Rekonstruktion des Bildes und dennoch werden diese Bereiche mit unendlich hohem PSNR bewertet.

Aufgrund dieses ausschlaggebenden Nachteils, der fehlenden Berücksichtigung auftretender Block- bzw. Bewegungsartefakte, ist das PSNR zur Qualitätsbewertung speziell von TPs nicht geeignet. Überdies handelt es sich beim PSNR um eine FR-Metrik und benötigt immer ein Originalsignal.

In dieser Arbeit soll nun eine Qualitätsmetrik untersucht werden, welche keine Referenz für die Bewertung benötigt. Ferner ist die Erkenntnis von Belang, inwieweit Block- bzw. Bewegungsartefakte bei der neuen NR-Metrik Berücksichtigung finden. Aus diesem Grund wird dieser referenzfreie Algorithmus des Natural Image Quality Evaluators hinsichtlich seiner Eignung zur Qualitätsbewertung verschiedener TPs aus der WT überprüft.

Im folgenden Kapitel 2 werden ein Einstieg zum Thema Qualitätsbeurteilung gegeben, die Berechnungsschritte der neuen Metrik aufgeführt und verschiedene Möglichkeiten präsentiert, das Modell zur Berechnung des NIQEs zu generieren. Kapitel 3 setzt sich aus der näheren Beschreibung der in dieser Arbeit verwendeten TPs zusammen. Das anschließende Kapitel 4 beschäftigt sich mit den Ergebnissen verschiedener Videosequenzen sowie Computertomographie-Daten (CT-Daten), welche mithilfe der in Kapitel 2 vorgestellten Modelle ausgewertet wurden. Im 5ten und letzten Kapitel erfolgen eine Zusammenfassung der erzielten Ergebnisse sowie ein Themenausblick.

Kapitel 2

Grundlagen Qualitätsbeurteilung

Dieses Kapitel basiert auf die Veröffentlichung des NIQEs. [MSB13]

2.1 Einführung zum Thema

Zunehmend an Bedeutung gewinnt der technische Fortschritt bei der Qualitätsbeurteilung von digitalen Bildern und Videos. Im März 2013 wurde diesbezüglich ein Modell von Anish Mittal, Rajiv Soundararajan und Alan C. Bovik veröffentlicht. Sie entwickelten den Algorithmus NIQE [MSB13], welcher die Qualität eines Bildes bestimmen kann, ohne vorherige Kenntnisse über das Bild oder dessen Verzerrungen zu benötigen. Bisherige NR-Modelle zur Qualitätsbeurteilung erforderten nähere Informationen zum zu bewertenden Bild. Diese wurden zum einen erlangt, indem die verschiedenen Modelle auf bestimmte Verzerrungsarten, wie z. B. Unschärfe oder Rauschen, trainiert wurden. Eine weitere Möglichkeit bestand darin, das Modell auf eine Datenbasis zu trainieren, welche sogenannte Human Opinion Scores enthält. Bei dem neuen Modell jedoch handelt es sich um einen Algorithmus, der beides nicht erfordert, daher auch der Name der Veröffentlichung *Completely Blind Image Quality Analyzer*. Darüber hinaus benötigt der NIQE auch keine Referenz, gegen welches das zu bewertende Bild verglichen werden soll. Bei dieser Methode wurde die Metrik lediglich auf eine Datenbasis, bestehend aus einer Reihe statistischer Merkmale von natürlichen und unverzerrten Bildern, trainiert. Aus diesem Grund ist der NIQE nur für natürliche Bilder ge-

dacht und findet für welche mit unnatürlichen Verzerrungen sowie Computergrafiken keine Anwendung.

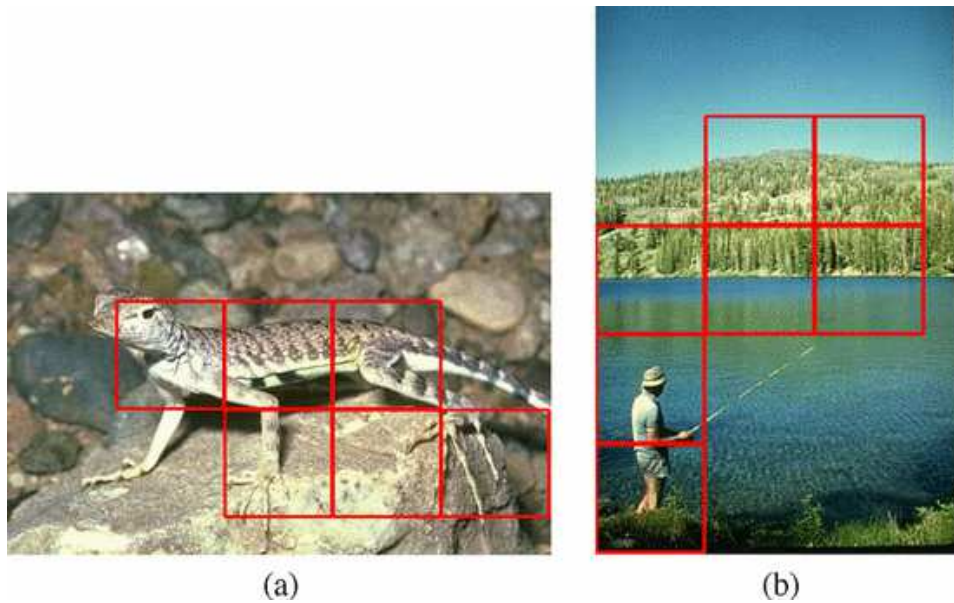
Im Grunde handelt es sich bei dieser Qualitätsmetrik um ein neues NR-Modell, bei dem Eigenschaften wie Opinion Unaware (OU) und Distortion Unaware (DU) erfüllt sind. OU und DU bedeuten wiederum, dass die Metrik weder eine subjektive noch eine objektive Bewertung benötigt und daher als OU-DU NR Image Quality Assessment (IQA) bezeichnet wird.

Den Erfindern zufolge eignet sich dieses Modell besser als das weit verbreitete PSNR für Videosequenzen (1) nach Formel $PSNR_V = 10 \times \log_{10}(255^2/MSE(image1, image2))$ bzw. für CT-Daten (2) nach Formel $PSNR_{CT} = 10 \times \log_{10}(4096^2/MSE(image1, image2))$. In der Praxis ist das PSNR, aufgrund dessen fehlender Berücksichtigung von Block- bzw. Bewegungsartefakten bei TPs mit Kompensationsverfahren, nicht sinnvoll einsetzbar. Da bei der Wavelet-Transformation mehrere Referenzbilder existieren, besteht zudem das Problem, dass generell nicht definiert ist, welches nun als Referenz zur Qualitätsbewertung herangezogen werden soll. Nach dem neuen IQA-Modell besteht zumindest dieses Problem nicht mehr. Wie sich die Block- bzw. Bewegungsartefakte von TPs mit Kompensationsverfahren auf den NIQE auswirken, soll ebenfalls in dieser Arbeit untersucht werden.

Mit dieser referenzfreien Metrik wurde ein Ansatz zur blinden Qualitätsbeurteilung entwickelt, welcher sich nach folgenden Berechnungsschritten ergibt.

2.2 Berechnung des NIQE-Algorithmus

Grundlage für die Berechnung bildet eine Zusammenstellung von Quality Aware Merkmalen, die aus einem Natural Scene Statistic (NSS) Modell gewonnen werden und zur Berechnung des NIQE-Indexes in ein multivariates Gauß Modell (MVG-Modell) eingesetzt werden. Somit ergibt sich als Definition der Qualität eines Bildes die Differenz zwischen einem MVG-Modell mit den NSS-Merkmalen vom Testbild und einem MVG-Modell mit den NSS-Merkmalen aus einer Datenbank von natürlichen Bildern.



[MSB13]

Abbildung 2.1: Patch Selection

2.2.1 Spatial Domain NSS

Die NIQE-Metrik basiert auf Merkmalen, welche natürlichen und unverzerrten Bildern zu entnehmen sind. Diese Merkmale werden nach folgendem Berechnungsschema aus lokalen Patches eines Bildes gewonnen.

Mithilfe der Formel $\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + 1}$ werden zunächst Koeffizienten (3) errechnet. Hierfür werden der Mittelwert (4) sowie die Varianz (5), definiert nach den Formeln $\mu(i, j) = \sum_{k=-3}^3 \sum_{l=-3}^3 w_{k,l} I(i+k, j+l)$ und $\sigma(i, j) = \sqrt{\sum_{k=-3}^3 \sum_{l=-3}^3 w_{k,l} [I(i+k, j+l) - \mu(i, j)]^2}$, herangezogen, die beide noch zusätzlich mit der Gaußfunktion bis zu einer Standardabweichung von ± 3 gewichtet werden.

2.2.2 Patch Selection

Als Nächstes wird das Bild, gemäß der Implementierung nach den Erfindern, in 96×96 große Patches geteilt, wie es in Abbildung (Abb.) 2.1 zu sehen ist. Da nahezu jedes Bild einen gewissen Grad an Verzerrung aufweist, werden nur diejenigen Patches gefiltert, welche die

meisten visuellen Bildinformationen besitzen. Um diese relevanten Patches zu finden, findet Formel $\delta(b) = \sum \sum_{(i,j) \in patch b} \sigma(i, j)$ Anwendung, mit welcher die lokale Schärfe (6) eines Bildes quantifiziert werden kann. Als Nächstes werden die Patches selektiert, die eine überdurchschnittliche Schärfe aufweisen, um ein Modell von Natural Image Patches zu konstruieren, um dann die NSS-Merkmale von den berechneten Koeffizienten \hat{I} eines jeden Patches zu ermitteln.

2.2.3 Characterizing Image Patches

Das resultierende Modell von Natural Image Patches wird anhand von Quality Aware NSS-Merkmalen charakterisiert, die von den selektierten Patches berechnet wurden. Dies erfolgt anhand der allgemeinen Gaußverteilung (7) nach Formel $f(x; \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp(-(\frac{|x|}{\beta})^\alpha)$, die das Verhalten der berechneten Koeffizienten \hat{I} , mit der Gamma Funktion (8) wie in Formel $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ definiert, erfasst.

Das Verhalten der allgemeinen Gaußverteilung $f(x; \alpha, \beta)$ mit den berechneten Koeffizienten \hat{I} wird jedoch negativ beeinflusst, sobald ein Bild Verzerrungen aufweist. Diese Abweichungen können ermittelt werden, indem Produkte benachbarter Koeffizienten \hat{I} in horizontaler, vertikaler und diagonaler Richtung gebildet werden. Für diese Produkte findet die Formel $f(x; \gamma, \beta_l, \beta_r) = \begin{cases} \frac{\gamma}{(\beta_l + \beta_r)\Gamma(\frac{1}{\gamma})} \exp(-(\frac{-x}{\beta_l})^\gamma) \forall x \leq 0 \\ \frac{\gamma}{(\beta_l + \beta_r)\Gamma(\frac{1}{\gamma})} \exp(-(\frac{x}{\beta_r})^\gamma) \forall x \geq 0 \end{cases}$ der asymmetrischen, standardisierten Gaußverteilung (9) Anwendung.

Aus dieser Berechnung ergeben sich 18 Parameter pro Dimension, demnach beträgt die Anzahl der Quality Aware NSS-Merkmale für Bilder 36.

2.2.4 Multivariate Gaussian Model

Die von Patches eines natürlichen Bildes berechneten NSS-Merkmale werden nun in das MVG-Modell (10) gemäß Formel $f_x(x_1, \dots, x_k) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \times \exp(-\frac{1}{2} (x - \nu)^T \Sigma^{-1} (x - \nu))$ eingesetzt.

2.2.5 NIQE-Index

Der NIQE-Index (11) resultiert aus der Gegenüberstellung des MVG-Modells für das Testbild und dem, welches anhand der Datenbank natürlicher Bilder erstellt wurde. Somit entspricht die Qualität eines Testbildes der Differenz zwischen einem MVG-Modell, errechnet aus 36 NSS-Merkmalen vom Testbild, und dem natürlichen MVG-Modell, errechnet aus NSS-Merkmalen der Datenbasis natürlicher Bilder nach Formel $D(\nu_1, \nu_2, \Sigma_1, \Sigma_2) = \sqrt{((\nu_1 - \nu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\nu_1 - \nu_2))}$. Je geringer der sich hieraus ergebende NIQE-Index ausfällt, umso besser ist die Qualität des zu bewertenden Bildes.

2.3 Modell Generierung für NIQE

Im Folgenden werden verschiedene Möglichkeiten zur Generierung eines NIQE-Modells präsentiert, die für diese Arbeit herangezogen werden. Beim ersten Modell, vorgestellt in Kapitel 2.3.1, handelt es sich um die Methode aus der Veröffentlichung [MSB13], die nur für natürliche Bilder Anwendung finden soll. In Kapitel 2.3.2 wird das NIQE-Modell bewusst auf verschiedene Arten von CT-Daten trainiert, um es dann auch auf medizinische Volumendaten, und damit nicht auf natürliche Bilder begrenzt, anwenden zu können. Das Ziel beim dritten Modell in Kapitel 2.3.3 ist die Metrik speziell auf eine aktuelle Videosequenz bzw. bestimmte Schicht von CT-Daten, also auf ein und dieselben Frames, zu trainieren und zu testen, wodurch letztendlich die NIQE-Indexe interpretiert werden sollen. Die jeweiligen Ergebnisse hierzu werden in Kapitel 4 diskutiert.

2.3.1 Modell aus Datenbank

Aus der Veröffentlichung [MSB13] geht hervor, dass das Modell auf NSS-Merkmalen basiert, welche aus einer Datenbasis natürlicher und unverzerrter Bilder abgeleitet wurden. Zur MVG-Modell Generierung, siehe Kapitel 2.2.4, wurden 125 natürliche Bilder mit Größen von 480×320 bis 1280×720 zugrunde gelegt. Die Bilder stammen nach Aussagen der Erfin-



[MSB13]

Figure 2.2: NIQE-Testbilder aus der Veröffentlichung

der aus Copyright Flickr Data und der Berkeley Image Segmentation Database. Dabei wurde darauf geachtet, dass sich die Bilder zur MVG-Modell Generierung mit dem Inhalt des Testbildes nicht überschneiden. Getestet wurde die Metrik auf fünf verschiedene Verzerrungsarten: JPEG/JPEG2000, additives weißes gaußsches Rauschen, Unschärfe und Rayleigh Fast Fading Channel. Aufgrund des Tests der Metrik bezüglich JPEG2000 wurde in dieser Arbeit auf eine Untersuchung der Skalierbarkeit bezüglich der Rate verzichtet. Unter Abb. 2.2 können Bilder mit unterschiedlichen Störungen betrachtet werden, die mit dem NIQE-Modell getestet wurden.

2.3.2 Modell aus CT-Datenbank

Bei diesem Modell wird die neue Metrik auf ganze Sequenzen verschiedener CT-Daten trainiert, was zum Ergebnis haben soll, dass die auf CT-Daten trainierte Metrik letztendlich auch für medizinische Volumendaten verwendet werden kann. Für Training und Test werden die Head-Sequenz, Thorax-Sequenz und die Neck-Sequenz herangezogen. Dieses Modell wird

jedoch nicht auf exakt gleiche Frames trainiert und getestet wie es in Kapitel 2.3.3 der Fall ist, stattdessen soll lediglich die Eignungsfähigkeit für CT-Daten auf die Probe gestellt werden.

2.3.3 Modell aus aktueller Sequenz bzw. bestimmter Schicht von CT-Daten

In diesem Kapitel wird die Metrik speziell auf eine aktuelle Videosequenz bzw. gezielt auf eine bestimmte Schicht von CT-Daten trainiert. Es handelt sich also bei Training und Test um identische Frames. Hierbei stellt sich jedoch die Frage, ob es dann noch als ein NR-Modell betrachtet werden kann. Die Metrik auf eine aktuelle Videosequenz oder bestimmte Schichten von CT-Daten zu trainieren, steht nämlich im Konflikt mit dem, was den NIQE eigentlich auszeichnet, denn die Besonderheit dieser Metrik liegt in der referenzfreien Qualitätsbeurteilung, selbst wenn diese Daten nur zur Generierung des Modells herangezogen werden. Anhand der Ergebnisse soll letzten Endes versucht werden die NIQE-Indexe zu interpretieren. Erwartungsgemäß müssten sich die NIQE-Indexe im Vergleich zu dem Modell nach Kapitel 2.3.1 erheblich verbessern bzw. sogar einen NIQE-Index von 0 erreichen, da auch das R-orig qualitativ beurteilt wird.

Kapitel 3

Berechnung der Tiefpassbänder aus der Wavelet-Transformation

Dieses Kapitel setzt sich aus einer Veröffentlichung einer verbesserten Methode zur Berechnung des L-FSEs vom Lehrstuhl für Multimediakommunikation und Signalverarbeitung (LMS) der Friedrich-Alexander-Universität Erlangen-Nürnberg zusammen. Weiterhin wird auf Teile des Kapitels 11 aus dem Vorlesungsskript 2012 der Veranstaltung Image and Video Compression des LMSs zurückgegriffen. [Kau12]

Von dauerndem Interesse in der Bild- und Videotechnik ist die Verringerung der Datenmenge, gleichzeitig soll sich aber auch die Qualität der digitalen Signale nicht verschlechtern. Ein wichtiger Schritt in Richtung skalierbarer Repräsentation des Originals wurde mit der WT gesetzt. Hierbei handelt es sich um eine Methode bei der das Originalsignal in einen niederfrequenten und einen hochfrequenten Anteil aufgeteilt wird. Während es sich bei den hochfrequenten Bildinformationen um die Kanten und Konturen eines Bildes handelt, werden die flächigen Strukturen im Tiefpass gefiltert. Das TP hat eine große Ähnlichkeit mit dem Originalsignal mit geringerer Auflösung. Damit es als herunterskalierte Version verwendet werden kann, wird eine hohe Qualität dessen benötigt.

Mithilfe der Erweiterung der WT um verschiedene Kompensationsverfahren kann eine deutliche Verbesserung der visuellen Qualität erzielt werden. Aufgrund der Entstehung von Arte-



Abbildung 3.1: Blockbasiertes Tiefpassband mit Artefakten

fakten im TP durch die Invertierung während der Transformation, wie sie in Abb. 3.1 zu sehen sind, wurde diesbezüglich ein Verfahren vom LMS entwickelt. Hierbei wurde die blockbasierte Kompensation um die FSE zur Reduzierung dieser Artefakte weiter ausgebaut.

Die Berechnung der TPs aus der WT gliedert sich in zwei Schritte. Im ersten Schritt, auch Prediction Step bezeichnet, wird zunächst das Hochpassband HP_t (12) eines Bildes ermittelt und anschließend, im Update Step, das TP LP_t (15). Dies erfolgt immerzu unter Verwendung eines Referenzbildes f_{2t-1} (14) und eines aktuellen Bildes f_{2t} (13) zu jedem Zeitpunkt t .

Daher werden zu Beginn alle Sequenzen geladen und für jeden Zeitpunkt t stets ein Referenzbild f_{2t-1} sowie ein Aktuelles f_{2t} bestimmt. Für die weitere Vorgehensweise werden, wie bei der Integer-Transformation üblich, bei den Videosequenzen, aufgrund ihrer 8 Bit Bilder, jeweils für die Referenzbilder f_{2t-1} sowie für die aktuellen Bilder f_{2t} zur vollständigen Invertierung die Gleitkommazahlen im Bereich von $[0 \ 1]$ auf einen Bereich von $[0 \ 255]$ Pixelwerte gemappt. [CDSY97]

Im Folgenden werden verschiedene TPs aufgeführt, welche in dieser Arbeit hinsichtlich zur Prüfung der Eignung des NIQEs herangezogen werden.

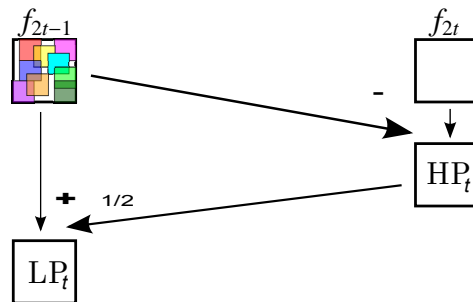


Abbildung 3.2: Blockschaltbild L-zero

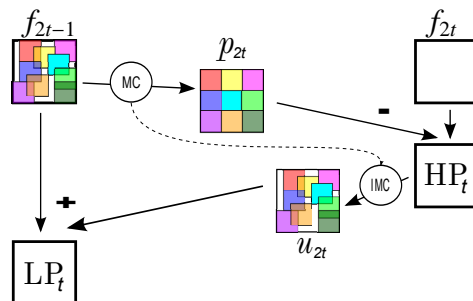


Abbildung 3.3: Blockschaltbild L-block

3.1 Tiefpassband ohne Kompensation

Bei der Berechnung des TPs aus der WT ohne Erweiterung um Kompensationsverfahren wird zunächst das Hochpassband HP_t im Prediction Step ermittelt, indem das jeweilige Referenzbild f_{2t-1} vom aktuellen Bild f_{2t} abgezogen wird, $HP_t = f_{2t} - f_{2t-1}$. Das TP LP_t ergibt sich dann im Update Step durch die Multiplikation des zuvor berechneten Hochpassbandes HP_t mit 0.5, Auf- bzw. Abrundung zum nächsten Integer und Addition zum Referenzbild f_{2t-1} , $LP_t = f_{2t-1} + \left\lfloor \frac{1}{2} \times HP_t \right\rfloor$, siehe Abb. 3.2.

Das hieraus resultierende TP entspricht zwar einer herunterskalierten Originalversion, stellt aber aufgrund der Mittelung aus zwei Frames ein unscharfes TP-Bild dar.

Im Folgenden wird das TP aus der WT ohne Kompensation als L-zero bezeichnet.

3.2 Tiefpassband mit blockbasierter Kompensation

Die Erweiterung um Kompensationsverfahren wird direkt in die Transformation mit eingebaut, siehe Abb. 3.3.

Hierbei wird das aktuelle Bild in Blöcke der Größe 16×16 Pixel geteilt und pro Block ein Bewegungsvektor geschätzt. Innerhalb eines Fensters von der hier festgelegten Größe von 15 Pixel wird nach der besten Übereinstimmung mit einem Block im Referenzbild zur Minimierung des Fehlers gesucht, dabei werden alle möglichen Positionen innerhalb dieses Suchbereichs getestet, bis das ideale Match gefunden ist. Zur Ermittlung des minimalen Fehlers wird die Summe der quadratischen Differenzen minimiert.

Aus dieser Berechnungsmethode ergibt sich der erforderliche Prädiktor p_{2t} (16). Auch hier wird zunächst das Hochpassband HP_t im Prediction Step berechnet, jedoch wird nun der eben ermittelte, aber gerundete Prädiktor p_{2t} , anstelle des Referenzbildes f_{2t-1} , von dem aktuellen Bild f_{2t} abgezogen, $HP_t = f_{2t} - \lfloor p_{2t} \rfloor$.

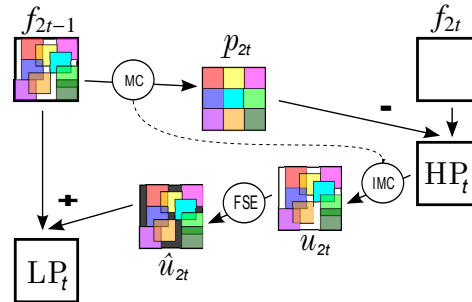
Als Nächstes folgt der Update Step zur Berechnung des TPs, in welchem die blockbasierte Kompensation erst einmal invertiert werden muss. Durch diese Invertierung verschieben sich die Pixel, wodurch im invers kompensierten Hochpassband drei verschiedene Arten von Verbundenheit dieser entstehen. Die Pixel können einfach, mehrfach oder gar nicht verbunden sein. Eine Methode zur Invertierung wurde von Herrn B. Girod und S. Han in der Veröffentlichung Optimum Update Operators for Lifting-Based Motion-Compensated Temporal Filtering vorgeschlagen, indem Gewichtungsfaktoren a_k (17) für mehrfach verbundene Pixel berechnet werden. Hierfür wird " $a_k = \frac{1}{k+1}$ "¹ für k-fach verbundene Pixel herangezogen und die nicht verbundenen Pixel wie bislang vom Referenzbild f_{2t-1} kopiert. Das sich daraus ergebende traditionelle Update Frame u_{2t} (18) mit gewichteten Pixeln wird zunächst gerundet und dann zum Referenzbild f_{2t-1} addiert, woraus sich letztendlich das TP LP_t ergibt, $LP_t = f_{2t-1} + \lfloor a_k \times u_{2t} \rfloor$.

Im Folgenden wird dieses TP aus der WT mit blockbasierter Kompensation als L-block bezeichnet.

¹[GH05]



Abbildung 3.4: unverbundene Pixel (weiße Stellen)



[SSK13]

Abbildung 3.5: Blockschaltbild L-FSE

3.3 Tiefpassband mit blockbasierter Kompensation und Lochfüllung

Die Berechnung des Hochpassbandes im Prediction Step entspricht exakt der Berechnung des Hochpassbandes aus der blockbasierten Kompensation. Der Unterschied dieser Methoden besteht lediglich im Update Step zur Berechnung des TPs.

Das Hauptaugenmerk wird hierbei auf die in Abb. 3.4 in weiß dargestellten, unverbundenen Pixel des invers kompensierten Hochpassbandes bzw. traditionellen Update Frames u_{2t} gesetzt. Diese unverbundenen Pixel, auch als Löcher bezeichnet, sollen nun mittels einer Lochfüllung rekonstruiert werden. Wie bereits erwähnt, entstehen diese Löcher durch die Invertierung der blockbasierten Kompensation im Update Step. In den Löchern sind keine Bildinformationen enthalten, wodurch sich eine Vermutung zur aktuellen Bewegung in diesen Bereichen schwierig gestaltet.

Aus diesem Grund wird hier auf die iterative Methode der FSE [SSK13] zurückgegriffen, mithilfe welcher verlorengegangene Bild- bzw. Videodaten rekonstruiert werden können. Das vollständige Schema ist in Abb. 3.5 dargestellt.

Für die Berechnung des invers kompensierten Hochpassbandes bzw. traditionellen Update

Frames u_{2t} wird hier die Mittelung mit einem Gewichtungsfaktor von $a_k = \frac{1}{k}$ durchgeführt. Zur Lochfüllung werden spezielle Parameter, wie sie vom LMS [SK10] festgesetzt wurden, übernommen. Nach diesem Berechnungsmodell, ergeben sich die rekonstruierten Pixel, dargestellt in schwarz unter Abb. 3.5 im rekonstruierten Update Frame \hat{u}_{2t} (19). Das rekonstruierte Update Frame \hat{u}_{2t} wird nun anstelle des traditionellen Update Frames u_{2t} zu dem Referenzbild f_{2t-1} addiert, $LP_t = f_{2t-1} + [\hat{u}_{2t}]$. Dies hat letztendlich das TP zum Ergebnis, welches im Folgenden als L-FSE bezeichnet wird.

Kapitel 4

Auswertung der Ergebnisse

In diesem Kapitel werden die Ergebnisse verschiedener Sequenzen, auf welche die zu untersuchende Metrik angewandt wurde, präsentiert.

4.1 Simulationssetup / Simulationsdaten

Zunächst sollen die Videosequenzen näher vorgestellt werden.

4.1.1 Basketball - Beschreibung

Bei der Basketball-Sequenz, zu sehen in Abb. 4.1, handelt es sich um ein Video mit sehr vielen und vor allem schnellen Bewegungen. Im Vordergrund ist ein kurzer Ausschnitt eines Basketballspiels mit stets wechselnder Position seitens der Spieler sowie des Balls zu sehen. Bei den Zuschauern im Hintergrund sind kaum Veränderungen auszumachen. Die Kamera wird von links nach rechts gedreht, dann eine bestimmte Szene herangezoomt und anschließend wieder weggerückt.

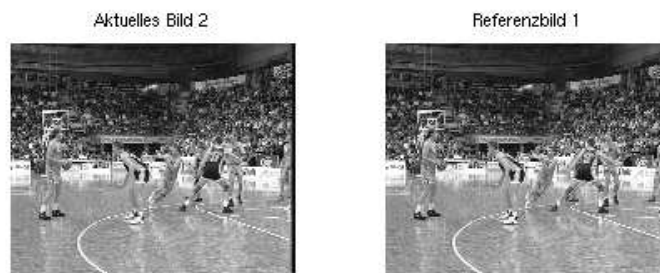


Abbildung 4.1: Ersten beiden Frames als Beispiel der Basketball-Sequenz

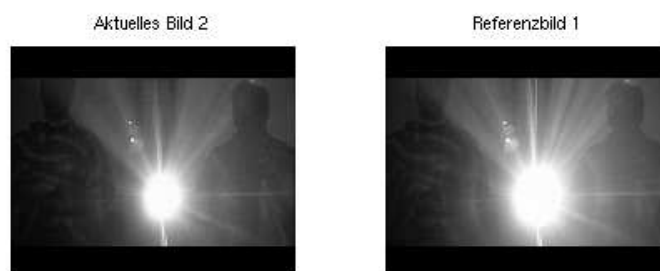


Abbildung 4.2: Ersten beiden Frames als Beispiel der Discovery-City-Sequenz



Abbildung 4.3: Ersten beiden Frames als Beispiel der Crew-Sequenz

4.1.2 Discovery-City - Beschreibung

Dieses Video enthält viele unterschiedliche Aufnahmen von Natur, Mensch und Stadt. Der Abb. 4.2 sind die ersten beiden Frames des ersten Kurzfilms zu entnehmen. Es weist außerdem viele Wechsel bei der Helligkeit auf und zeigt Schnellaufnahmen vom Straßenverkehr. Generell sind die Bewegungen sehr schnell, das Besondere bei der Discovery-City-Sequenz jedoch sind die Transitionen. An diesen Übergängen zwischen zwei verschiedenen Aufnahmen sind des Öfteren Bilder zu sehen, welche aus mehreren Frames, den letzten der vorherigen und den ersten der folgenden Aufnahme, bestehen.

4.1.3 Crew - Beschreibung

Dieses Video beinhaltet eine Gruppe von Astronauten, die sich auf die Kamera zu bewegen, wie in Abb. 4.3 dargestellt. Bedeutungsvoll in diesem Kurzfilm sind die Blitzlichter, anhand welcher sich die Helligkeit von einem Frame zum nächsten teilweise erheblich ändert. Weiterhin bilden sich unter Einfluss dieser sich stets verändernde Schatten.



Abbildung 4.4: Ersten beiden Frames als Beispiel der Flower-Sequenz

4.1.4 Flower - Beschreibung

Bei diesem Video sind, vgl. Abb. 4.4, Häuser im Hintergrund und ein Blumenbeet mit einem Baum im Vordergrund zu sehen. Alles was sich hier bewegt ist die Kamera selbst, diese wird von links nach rechts gefahren.

4.1.5 Foreman - Beschreibung

Die Foreman-Sequenz zeigt zunächst, in Abb. 4.5 einzusehen, einen Bauarbeiter und unmittelbar hinter ihm Steinblöcke eines Krans. Als der Mann beginnt zu sprechen bewegen sich sein Kopf und seine Gesichtszüge. Im letzten Drittel des Videos wird die Kamera plötzlich von links nach rechts geschwenkt und dann wieder ruhig gehalten.

4.1.6 Discovery-Orient - Beschreibung

Dieses Video, die ersten beiden Frames sind in Abb. 4.6 abgebildet, besteht aus vielen kurzen und unterschiedlichen Aufnahmen. Zwischen den verschiedenen Kurzfilmen sind teil-

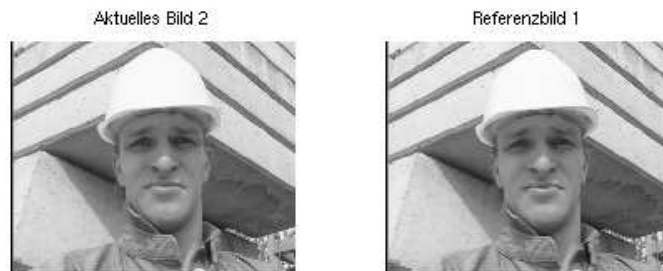


Abbildung 4.5: Ersten beiden Frames als Beispiel der Foreman-Sequenz



Abbildung 4.6: Ersten beiden Frames als Beispiel der Discovery-Orient-Sequenz

weise starke Helligkeitsunterschiede auszumachen, darüber hinaus bewegt sich die Kamera durchgehend gemächlich und es sind gelegentlich Bewegungen der Bildobjekte vorhanden. Transitionen, wie sie bei der Discovery-City-Sequenz mit mehreren Frames ineinander zu sehen sind, können hier nicht vernommen werden.

4.2 NIQE mit Datenbank-Modell im Vergleich mit PSNR

In den Kapiteln 4.2.1 bis 4.2.6 werden die Ergebnisse der NIQE-Metrik mit der Modellrechnung aus einer Datenbank, wie in Kapitel 2.3.1 beschrieben, diskutiert und auf die unter Kapitel 4.1 beschriebenen Videosequenzen angewandt. Kapitel 4.2.7 stellt einen Versuch dar, die NIQE-Metrik mit demselben Modell auf unnatürliche Bilder anzuwenden.

In allen aufgezeigten Plots werden jeweils die gleichen Farben für die verschiedenen TPs verwendet, demnach entspricht rot dem L-zero, blau dem L-block, cyan dem L-FSE und magenta stellt das pro TP aktuelle, originale Referenzbild f_{2t-1} R-orig dar. Unter dem Begriff Originalvideo, jeweils aufgegriffen in der Auswertung der Mittelwerte, ist die gesamte Sequenz mit allen vorhandenen Frames zu verstehen.

4.2.1 Basketball

Die wesentlichen Eigenschaften der Basketball-Sequenz sind in Kapitel 4.1.1 beschrieben.

4.2.1.1 NIQE

Die Untersuchung des L-zeros zu Beginn des Videos ergibt keine verwertbaren Ergebnisse der neuen Metrik. Bei näherer Betrachtung können zwar geringfügige Unterschiede im Video festgestellt werden, jedoch lediglich im Zusammenhang mit der gefilmten Szene. Die Qualität für die ersten zehn L-zeros scheint optisch gleich, trotzdem existieren bei der Auswertung unterschiedliche NIQE-Indexe im Bereich von ca. 6–8. Im Vergleich hierzu fallen die NIQE-Indexe für das L-block und L-FSE deutlich niedriger aus. Die TP-Bilder sind zwar nicht so verschwommen, jedoch sind neben einigen unscharfen Stellen mehrere Artefakte vorhanden

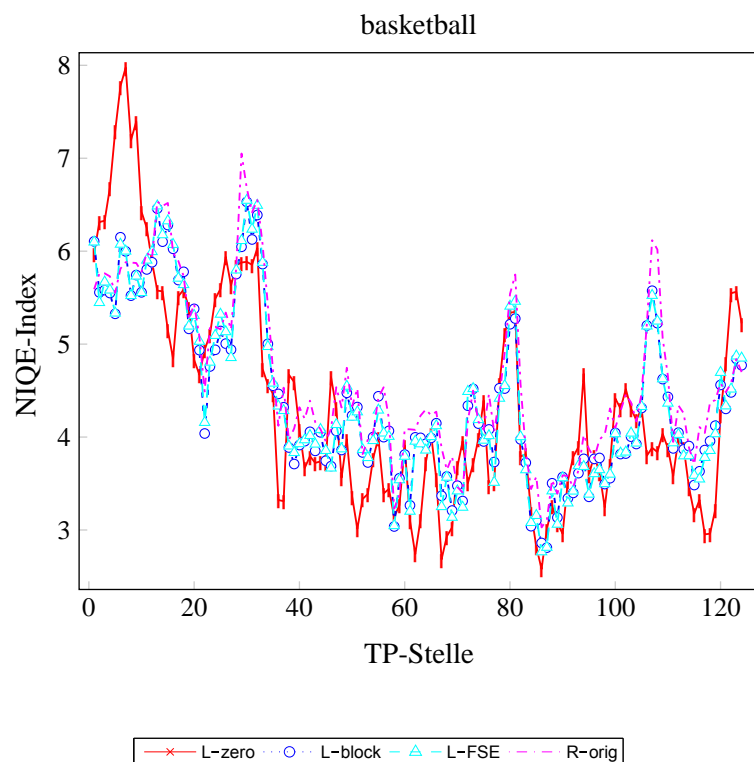


Abbildung 4.7: Ergebnisse der NIQE-Metrik von der Basketball-Sequenz

und dennoch sind die NIQE-Indexe stets besser. Die NIQE-Indexe für L-block und L-FSE unterscheiden sich kaum, aber auch hier gehen leichte Schwankungen im Kurvenverlauf hervor. Bei der Auswertung des R-orig innerhalb der ersten zehn TP-Stellen tritt das Auf und Ab der NIQE-Indexe nicht auf. Diese sind hier nahezu identisch, aber fast durchgehend schlechter als die von L-block bzw. L-FSE, trotz bester visueller Qualität.

Beachtlich ist die TP-Stelle 29, die aus den Frames 57, hierbei dem Referenzbild f_{2t-1} R-orig, und 58, dem aktuellen Bild f_{2t} , gebildet wird. Das L-zero ist so unscharf abgebildet, dass einzelne Spieler nicht mehr zu erkennen sind, allerdings ist der NIQE-Index am niedrigsten, was der besten Qualität dieser TP-Stelle entspricht. Bei L-block sowie L-FSE ist die Szene trotz Artefakten viel deutlicher, dennoch erzielen diese einen schlechteren NIQE-Index. Vergleicht man die TP-Bilder mit dem R-orig, sind Unterschiede bei näherer Betrachtung wahrnehmbar, es wird aber erneut, ungeachtet bester Optik, am schlechtesten bewertet.

Bei den TP-Stellen 61 und 62 ist ebenso ein interessantes Ergebnis auszumachen, wie der

Abb. 4.7 zu entnehmen ist. Obwohl kaum ein visueller Qualitätsunterschied erkennbar ist, verbessert sich der NIQE-Index um 0.36 auf 2.74 von L-zero, im Gegensatz zu L-block und L-FSE, wo sich die NIQE-Indexe von 3.2 auf nahezu 4 wieder erhöhen. Ein Blick auf die jeweiligen R-origs zeigt, dass die NIQE-Indexe hierbei annähernd gleich bleiben. Demnach ergeben sich bei diesen TPs sehr unterschiedliche und zusammenhangslose Ergebnisse.

Ab der TP-Stelle 36 verkleinern sich die NIQE-Indexe für alle TPs spürbar, dennoch bleibt L-zero stets unscharf. L-block sowie das L-FSE weisen mit ähnlichen NIQE-Indexen durchgehend Artefakte auf. Sie sind zwar schärfer dargestellt als L-zero, werden aber oft unerklärbar schlechter bewertet.

Die Analyse des besten bzw. schlechtesten NIQE-Indexes der gesamten Basketball-Sequenz führt ebenso zu keinem eindeutig interpretierbaren Ergebnis. Der niedrigste NIQE-Index von 2.58 wird von L-zero an der TP-Stelle 86 erzielt. Die NIQE-Indexe dieser TPs fallen generell sehr klein aus, trotzdem ist der des R-origs am höchsten. Anders als beim schlechtesten NIQE-Index der Sequenz i. H. v. von 7.976 an der TP-Stelle 7, der ebenfalls von L-zero erreicht wurde. Auch hier liegen alle NIQE-Indexe sehr hoch, jedoch hier mit dem R-orig als Bild mit der besten Qualität.

Die Auswertung des L-zeros lässt keine eindeutigen Aussagen über diese neue Metrik zu. Das L-zero ist während der gesamten Sequenz durchgehend verschwommen und es wird zunehmend unschärfer, sobald Bewegungen näher an der Kamera stattfinden. Ein Blick auf die NIQE-Indexe zeigt jedoch, dass sich diese, mit leichter Tendenz zum Besseren, verringern. Weiterhin auffallend ist, dass die Ergebnisse bei aufeinanderfolgenden TPs teilweise starken Schwankungen unterliegen, vgl. rote Kurve an den TP-Stellen 32 und 33 in Abb. 4.7, wobei sich die visuelle Bildqualität kaum verändert.

Beim Vergleich des L-blocks mit dem L-FSE ist kaum ein optischer Unterschied wahrnehmbar, die NIQE-Indexe der kompensierten TPs liegen auch stets auf nahezu gleicher Höhe. Bei näherer Betrachtung einzelner Bildinformationen fällt auf, dass diese aufgrund vieler und vor allem schneller Bewegungen nur schlecht rekonstruiert werden können. Zusätzlich zu den schnellen Bewegungen hat auch die Kameradrehung bzw. das Zoom-In oder Zoom-Out einer bestimmten Szene einen negativen Einfluss auf die Bildqualität. So sind u. a. vermehrt

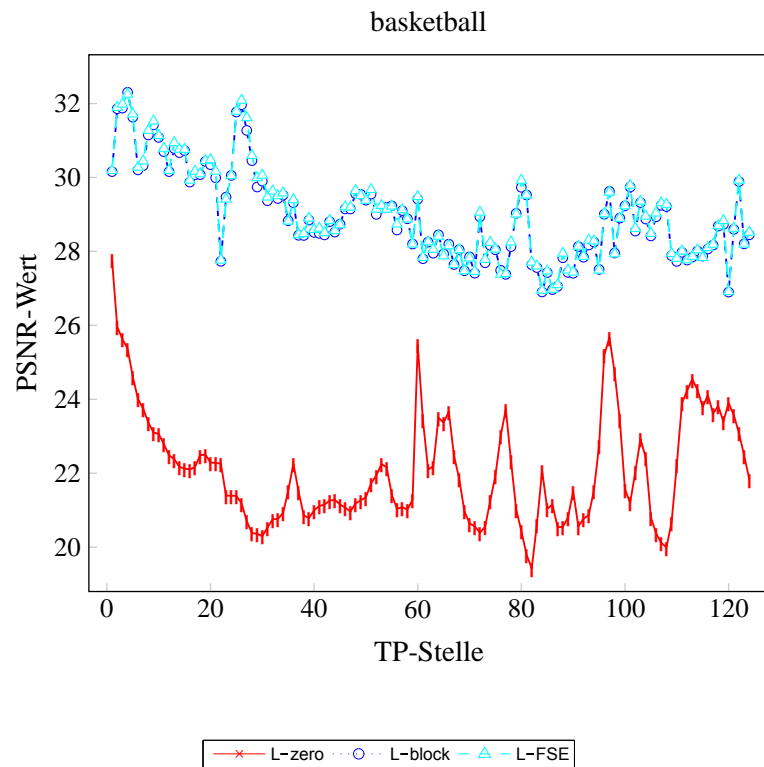


Abbildung 4.8: Ergebnisse des PSNRs von der Basketball-Sequenz

Blockartefakte an Armen und Beinen der Spieler oder am Ball auszumachen. Dies spiegelt sich allerdings nicht in den NIQE-Indexen wieder. Beispielsweise ist in den TP-Bildern der TP-Stellen 5 und 6 lediglich eine Bewegung des Fußes eines Spielers zu erkennen, dennoch werden die NIQE-Indexe um ca. 0.7 schlechter. Eindeutig wahrzunehmen ist die Vermehrung von Blockartefakten einer Szene der TP-Stellen 15 und 16, der NIQE-Index aber verkleinert sich. Dies ist in umgekehrter Weise bei den TP-Stellen 61 und 62 zu sehen, wo sich die Artefakte leicht verringern, der NIQE-Index jedoch steigt.

Was den Einfluss der Bildschärfe und der Blockartefakte auf den NIQE angeht, kann bislang noch keine präzise Aussage getroffen werden. Aus diesem Grund soll die Metrik noch auf andere Sequenzen angewandt werden.

Basketball	L-zero	L-block	L-FSE	R-orig
Mittelwert NIQE als Index	4.362	4.418	4.421	4.638
Mittelwert PSNR in dB	22.11	28.98	29.03	infinity

Tabelle 4.1: Mittelwerte Basketball-Sequenz

4.2.1.2 PSNR

Die Auswertung der Basketball-Sequenz mithilfe des PSNRs liefert ein eindeutig interpretierbares Ergebnis. Nahezu durchgehend erzielt L-FSE mit wenigen Ausnahmen die höchsten PSNR-Werte. Mit nicht allzu viel schlechteren PSNR-Werten folgen die Ergebnisse für L-block und an letzter Stelle steht erwartungsgemäß das L-zero. Die PSNR-Werte für L-zero liegen während der gesamten Sequenz um durchschnittlich ca. 7 dB niedriger.

Wie bereits erwähnt, ist ein leichter Rückgang der Bildqualität dieses Videos wahrnehmbar, je näher sich die Szene vor der Kamera abspielt. Dies korrespondiert auch mit den PSNR-Werten für alle TPs, wo eine leichte Tendenz zum Schlechteren zu erkennen ist, was gut in Abb. 4.8 zu vernehmen ist.

Die untersuchten Extremwerte, welche sich bei der Auswertung mit der NIQE-Metrik ergeben, sind nicht zwingend auch beim PSNR herausragend. Diese fehlende Übereinstimmung kann mit einem Vergleich beider Metriken bezüglich des besten bzw. schlechtesten NIQE-Indexes näher gebracht werden. Gemäß dem NIQE hat das L-zero an der TP-Stelle 86 die beste Bildqualität der gesamten Sequenz. Die Auswertung mit dem PSNR zeigt jedoch, dass sich hier sogar ein eher schlechter PSNR-Wert für alle TPs ergibt. Ebenso widersprüchlich erweist sich die Betrachtung der TP-Stelle 7 mit der schlechtesten Qualität, welches mithilfe des PSNRs wiederum als sehr gut bewertet wird. Darüber hinaus ist der PSNR-Wert höher als der an der TP-Stelle 86 ermittelte. Die Ergebnisse der Metriken sind hierzu gegensätzlich.

4.2.1.3 NIQE vs. PSNR

4.2.1.3.1 Betrachtung der Mittelwerte Als Nächstes sollen die Mittelwerte nach Tabelle (Tab.) 4.1 betrachtet werden. Demnach ergibt sich der niedrigste Mittelwert der gesamten

Sequenz beim NIQE aus dem L-zero, während die Mittelwerte für L-block und L-FSE quasi identisch sind und 4.4 betragen. Weshalb das R-orig am schlechtesten ausfällt, kann nicht begründet werden. Die Auswertung des Originalvideos mit 250 Frames ergibt im Vergleich zu R-orig keine nennenswerte Modifizierung. Das Resultat der Mittelwerte stellt ein Indiz dafür dar, dass die Qualitätsbewertung mithilfe dieser Metrik keine verlässlichen Ergebnisse liefert.

Anders verhalten sich die Mittelwerte nach dem PSNR, dessen Ergebnisse dem Erwartungsfall entsprechen. Das R-orig liegt im unendlichen Bereich, an zweiter Stelle steht das L-FSE, dicht gefolgt von L-block und am schlechtesten kommt das L-zero mit dem niedrigsten PSNR-Wert weg. Da Block- bzw. Bewegungsartefakte beim PSNR keine Berücksichtigung finden, welche im L-block bzw. L-FSE auftreten, ergeben sich für die beiden kompensierten TPs stets höhere PSNR-Werte als für L-zero.

4.2.1.3.2 Betrachtung der Plots Vergleicht man die Plots beider Qualitätsmetriken, siehe Abb. 4.7 und Abb. 4.8, zunächst für das L-zero, wird schnell klar, dass hierbei keine Übereinstimmungen zu finden sind. Der Verlauf der Kurven für L-zero ist äußerst unterschiedlich, demnach korrespondieren auch die Extremwerte bei NIQE und PSNR nicht miteinander. Weiterhin auffallend ist, dass die Qualitätsbeurteilung mithilfe des NIQEs deutlich stärkeren Schwankungen unterlegen ist als beim PSNR, selbst dann, wenn sich das Bild nicht wesentlich verändert.

Ebenso unerklärlich sind die Kurvenverläufe für L-block bzw. L-FSE, die bei beiden Metriken nahezu übereinander liegen. Beim Kurvenverlauf des PSNRs sowie des NIQEs ist jeweils keine Ähnlichkeit zu erkennen. Daher sind auch keine gemeinsamen Extremwerte beider Metriken auszumachen. Beispielsweise ergibt sich an der TP-Stelle 22 anhand des NIQEs eine gute Qualität. Wertet man dieselbe TP-Stelle mit dem PSNR aus, so tritt genau das Gegenteil ein. Des Weiteren ist beim PSNR im Laufe der Sequenz eine leichte Tendenz zur Qualitätsverschlechterung wahrnehmbar, was durchaus mit der Szene des Basketballspiels, welche im fortgeschrittenen Video näher an der Kamera stattfindet, begründet werden kann. Im Gegensatz hierzu wendet sich der NIQE leicht dem Besseren zu. Ähnlich wie beim

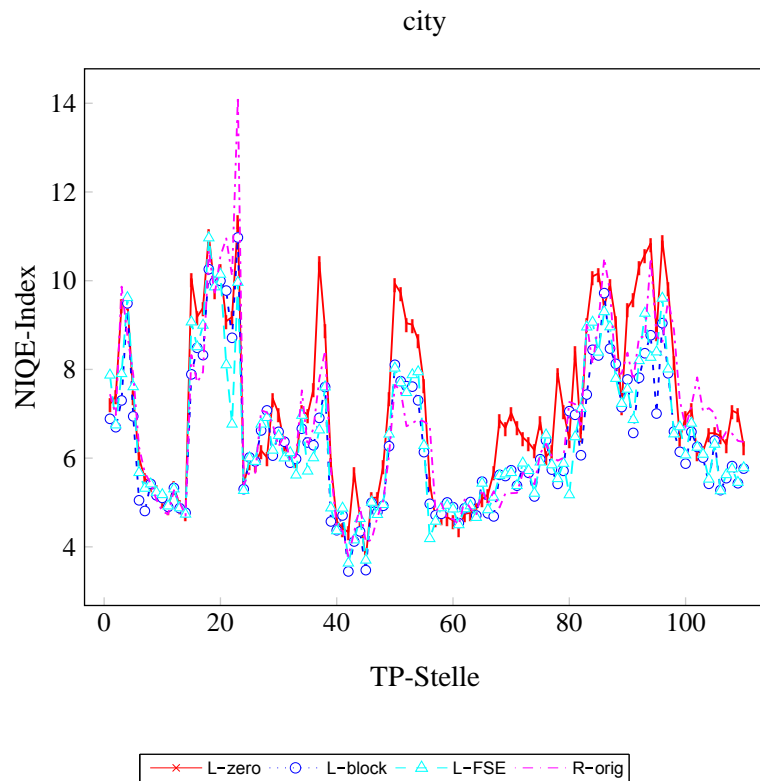


Abbildung 4.9: Ergebnisse der NIQE-Metrik von der Discovery-City-Sequenz

L-zero weisen auch hier die Kurvenverläufe beider kompensierter TPs starke Schwankungen auf.

Das PSNR für das R-orig befindet sich wie erwartet im unendlichen Bereich. Beim NIQE jedoch liegt diese fast auf gleicher Höhe mit den anderen TPs, was nicht gerade für diese neue Metrik spricht, da man für die jeweiligen originalen Referenzbilder eine deutlich höhere Qualität bzw. den NIQE-Index näher an 0 erwarten würde.

4.2.2 Discovery-City

Die wesentlichen Eigenschaften der Discovery-City-Sequenz sind in Kapitel 4.1.2 beschrieben.

4.2.2.1 NIQE

Die Untersuchung der NIQE-Indexe bei der City-Sequenz gestaltet sich nicht einfach, da aufgrund der Zusammenstellung vieler Kurzvideos die Ergebnisse untereinander nur schlecht verglichen werden können.

Gleich an der ersten TP-Stelle weichen die NIQE-Indexe etwas voneinander ab. Das L-zero, welches hier optisch dem R-orig gleicht, stellt das qualitativ beste TP-Bild dar, beispielsweise können Bildinformationen wie zwei Lampen und die Krawattenschleife entnommen werden, zur Veranschaulichung der ersten beiden Frames siehe erneut Abb. 4.2. Dies wird bei L-block und L-FSE schlecht rekonstruiert, u. a. sind die Lampen mehrfach abgebildet. Weiter sind beide kompensierten TP-Bilder sehr vielen Blockartefakten unterlegen, jedoch erzielt L-block mit den meisten Artefakten den kleinsten NIQE-Index von 6.886 an dieser TP-Stelle, während L-FSE mit 7.872 am schlechtesten ausfällt.

Wie in Abb. 4.9 zu sehen ist, werden die NIQE-Indexe für alle TPs von der TP-Stelle 14 auf 15 abrupt um ca. 5 schlechter. Der Grund könnte sein, dass an diesen TP-Stellen ein Kurzvideo endet und das nächste beginnt. Für dieses folgende Video ab TP-Stelle 15, in welchem viele kleine Lämpchen zu sehen sind, werden die NIQE-Indexe generell für alle TPs schlechter. Unterschiede bei den einzelnen TP-Bildern sind während dieses Kurzfilms kaum welche auszumachen, dennoch bewegen sich die NIQE-Indexe zwischen 7.8 und 10. Die TP-Bilder sind hier immerzu leicht verschwommen, obwohl lediglich die Helligkeit eine geringfügige Änderung erfährt. Der Grund weshalb sich die NIQE-Indexe so plötzlich und extrem verschlechtern ist nicht eindeutig ersichtlich.

Äußerst interessant ist der NIQE-Indexsprung von der TP-Stelle 22 auf 23, an welchem die Kurzaufnahme schnelle und gravierende Helligkeitswechsel enthält. Die Rangfolge der TPs bleibt gleich. Demnach erzielt L-FSE die niedrigsten NIQE-Indexe gefolgt von L-block. Beide weisen TP-Bilder mit einem sehr hohen Anteil an Blockartefakten auf, wessen Rekonstruktion bei L-FSE auch teilweise gelungen ist und dennoch ist die visuelle Qualität sehr schlecht. Beim L-zero reagiert die Metrik anders als erwartet. In diesen TP-Bildern wird der Kontrastunterschied zwei aufeinanderfolgender Frames sehr weich gestaltet. Hierbei sind

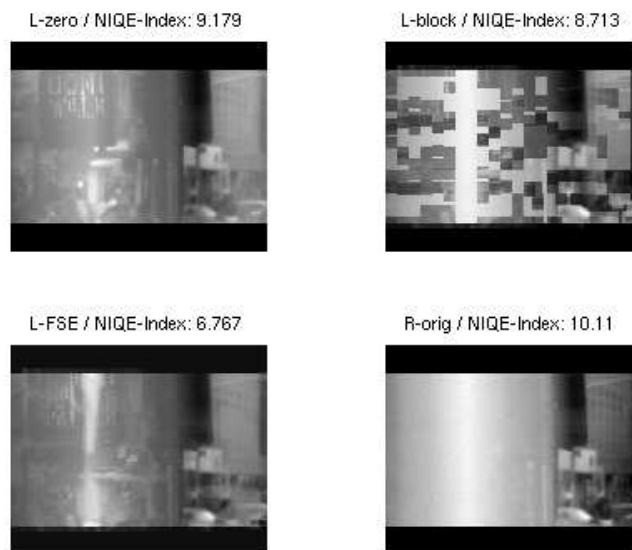


Abbildung 4.10: TP-Bilder der TP-Stelle 22 der Discovery-City-Sequenz

innerhalb eines TP-Bildes zwei Frames abgebildet, aber man kann beide deutlich erkennen, vgl. TP-Stelle 22 in Abb. 4.10. Der NIQE jedoch bewertet das L-zero äußerst schlecht. Die R-origs, welche jeweils lediglich die zugehörigen Referenzbilder zeigen, beurteilt die Metrik unergründlich am schlechtesten.

Die folgende TP-Stelle 37 wird wiederum auffallend schlecht bewertet und die nähere Betrachtung dieses Kurzfilms zeigt, dass es erneut eine kurze Szene im Schnelldurchlauf wiedergibt. Aus diesem Grund werden alle TP-Bilder unscharf dargestellt. L-block weist viele Artefakte auf, die bei L-FSE, neben vielen unscharfen Stellen, ebenso vorhanden sind. Diese beiden TPs erzielen trotzdem die besten NIQE-Indexe an dieser TP-Stelle. Für das optisch bessere L-zero und das R-orig wiederum liegen die NIQE-Indexe unbegründet deutlich höher.

Als Nächstes wird erneut ein Übergang zwischen zwei Kurzaufnahmen an der TP-Stelle 56 untersucht. Die Rangfolge der TPs entspricht derselben wie an den TP-Stellen 22 und 23. Demnach zeigt sich optisch das L-zero am besten mit zwei deutlich erkennbaren Frames, erzielt aber den hierbei höchsten NIQE-Index von 5.448. L-block besitzt einen großen An-

teil an Artefakten und L-FSE mit einem NIQE-Index von 4.186 wiederum kann diese, durch das Auftreten von zwei völlig verschiedenen Frames, nur teilweise rekonstruieren. Dass das L-zero trotz bester visueller Qualität am schlechtesten wekommt, kann auch hier nicht erklärt werden.

Die TP-Bilder von der TP-Stelle 42 auf 43, welches lediglich Äste und Sonnenstrahlen aufweist und Bewegungen kaum vorhanden sind, scheinen optisch gleich. Dennoch enthält diese Szene einen NIQE-Indexsprung zum Schlechteren i. H. v. ca. 1. Generell sind die NIQE-Indexe innerhalb dieser TP-Stellen im Vergleich zur gesamten Sequenz am niedrigsten.

Darüber hinaus befindet sich das qualitativ hochwertigste TP-Bild der Sequenz von L-FSE an der TP-Stelle 42, ohne erkennbare Artefakte, mit einem NIQE-Index von 3.446. Warum sich allerdings die NIQE-Indexe von einer TP-Stelle zur nächsten um 1 erhöhen obwohl die TP-Bilder optisch gleich bleiben, bleibt offen. Bei dem vom NIQE als schlechtestes bewertetes Bild der Discovery-City-Sequenz handelt es sich unerklärlich um das R-orig an der bereits untersuchten TP-Stelle 23.

Wie schon erwähnt, können nicht alle TPs untereinander, aufgrund verschiedener Kurzvideos, verglichen werden, dennoch können zu einigen Ergebnissen Aussagen getroffen werden. Zum Beispiel stechen die L-zeros hervor, die sich aus den Transitionen ergeben. Diese stellen an den untersuchten TP-Stellen optisch gesehen die besten TP-Bilder dar, da hierbei die Übergänge eher weich gestaltet sind und dadurch beide Frames, sozusagen das Letzte aus dem vorhergehenden Video und dem Ersten aus der folgenden Aufnahme, sehr gut und deutlich erkennbar sind. Dennoch erzielt L-zero hierbei nahezu durchgehend den höchsten NIQE-Index. Da dieses Video generell eher Schnellaufnahmen beinhaltet, ist L-zero an den untersuchten TP-Stellen grundsätzlich verschwommen und am schlechtesten.

Bei L-block und L-FSE können die Ergebnisse gemeinsam betrachtet werden. Bei den Transitionen sind erheblich viele Blockartefakte beim L-block auszumachen, diese wirken umso gravierender je größer der Kontrast beider Frames ist. Diese Artefakte sind, etwas dezent, auch bei L-FSE auffindbar, welches aber noch zusätzlich viel unschärfer abgebildet ist. Dennoch sind die NIQE-Indexe bei L-FSE an den Übergängen niedriger als bei L-block und das teilweise zwischen 0.8–2. Ebenso unstimmg sind die Resultate beider TP-Bilder, die

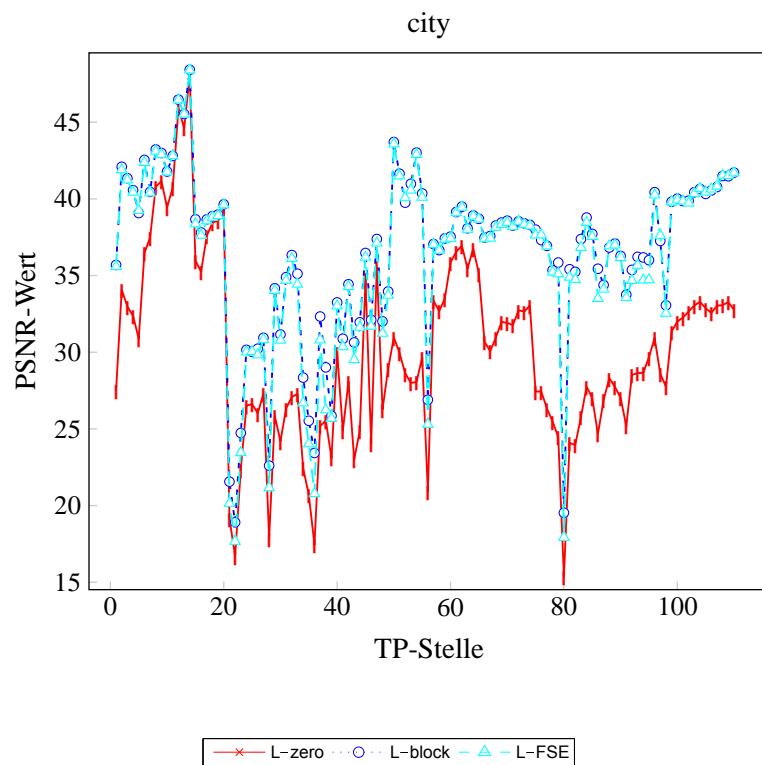


Abbildung 4.11: Ergebnisse des PSNRs von der Discovery-City-Sequenz

aufgrund vieler und schneller Bewegungen teilweise stark verschwommen sind und die Störungen der TP-Bilder denen bei den Transitionen entsprechen. Dennoch wechseln sich die NIQE-Indexe beider TPs ab, sodass eine plausible Erklärung ausbleibt.

Was die NIQE-Indexe beim R-orig betrifft, fallen diese stets zu hoch aus. Dass es sich hierbei um das originale Referenzbild f_{2t-1} handelt, scheint die Metrik nicht positiv zu beeinflussen.

4.2.2.2 PSNR

Wertet man nun dieselbe Sequenz mithilfe des PSNRs, siehe Abb. 4.11, aus, so ist ähnlich wie bei der Basketball-Sequenz eine Rangfolge erkennbar. Die PSNR-Werte für das R-orig befinden sich wie erwartet im unendlichen Bereich, danach folgt das L-block mit nicht sehr viel höheren Ergebnissen als beim L-FSE. Das L-zero erzielt für die gesamte Sequenz die schlechtesten PSNR-Werte und liegt stets zwischen 1 und 10 dB unter den kompensierten TPs.

Discovery-City	L-zero	L-block	L-FSE	R-orig
Mittelwert NIQE als Index	7.127	6.381	6.448	6.772
Mittelwert PSNR in dB	30.12	36.45	36.1	infinity

Tabelle 4.2: Mittelwerte Discovery-City-Sequenz

Die Transitionen haben auch hier einen negativen Einfluss auf das PSNR. Bei den untersuchten Übergängen liegen die Ergebnisse weit unter dem Durchschnitt. Hier ist deutlich zu erkennen, dass diese Qualitätsmetrik nicht sensibel auf Blockartefakte reagiert, da stets die besten Ergebnisse von L-block erzielt werden, gefolgt von dem nicht allzu viel schlechter ausfallendem L-FSE, obwohl diese TP-Bilder starke Artefakte aufweisen. Im Gegensatz zum L-zero, welche keine enthalten, aber bei den Transitionen aufgrund schneller Bewegungen stellenweise verschwommen ist und oft als Schlechtestes wekommt. In diesem Punkt stimmen NIQE und PSNR überein, sie bewerten die Qualität eines TP-Bildes, welches aus zwei Frames ineinander besteht, in diesem Fall L-zero, unterdurchschnittlich schlecht.

Bei der Untersuchung des vom NIQE als qualitativ bestes bzw. schlechtestes TP-Bild decken sich die Ergebnisse beider Metriken nicht. Demnach schneidet das gemäß dem NIQE am besten bewertete TP-Bild beim PSNR eher unterdurchschnittlich ab. Die PSNR-Werte für das als schlechtestes geltendes TP-Bild befinden sich zwar auch beim PSNR unter dem Durchschnitt, jedoch gehört es nicht zu den herausragend schlechtesten dieser Sequenz.

4.2.2.3 NIQE vs. PSNR

4.2.2.3.1 Betrachtung der Mittelwerte Stellt man die Mittelwerte beider Metriken gegenüber und lässt man zunächst das R-orig außen vor, so stimmt die Rangfolge überein, zu sehen in Tab. 4.2. Auffallend ist, dass auch hier L-zero beim NIQE das qualitativ minderwertigste TP darstellt und L-block sowie L-FSE wiederum nur geringfügig voneinander abweichen. Das R-orig liegt mit einem NIQE-Index von 6.772 dazwischen. Durch die Anwendung des NIQEs auf das Originalvideo mit 220 Frames ergibt sich ein NIQE-Index von 6.672, es ist also nur eine kleine Verbesserung im Vergleich zu R-orig auszumachen und

an der Rangfolge ändert sich nichts. Warum die originalen Referenzbilder f_{2t-1} nur durchschnittlich ausfallen kann nicht begründet werden.

Beim PSNR fallen die Ergebnisse wie erwartet aus. Demnach und mit einer nur geringfügigen Unterscheidung zwischen L-block und L-FSE liegen diese TPs mit ihren PSNR-Werten um durchschnittlich ca. 6 dB höher als das L-zero. Auch hier befindet sich das R-orig im unendlichen Bereich.

4.2.2.3.2 Betrachtung der Plots Schenkt man den Plots beider Metriken Beachtung, siehe dazu Abb. 4.9 und Abb. 4.11, so kann auch diesbezüglich keine präzise Aussage gefällt werden. Beim Vergleich des L-zeros passt weder der Kurvenverlauf zusammen, noch korrespondieren die Extremwerte miteinander und eine Übereinstimmung für dieses TP bleibt daher aus. Die Auswertung liefert teilweise sogar gegensätzliche Ergebnisse. Beim PSNR befindet sich die Kurve für L-zero durchgehend unterhalb der anderen beiden TPs, was man anhand der Bewertung mit dem NIQE wiederum nicht behaupten kann, denn hierbei liegt L-zero oft auf gleicher Höhe mit L-block bzw. L-FSE.

Beim PSNR liegen die Kurven für L-block sowie L-FSE nahezu übereinander und weisen dieselben Extremwerte wie die von L-zero auf, daher sind auch hier keine entsprechenden Ergebnisse zwischen PSNR und NIQE erkennbar. Ebenso wenig ähnelt sich der Kurvenverlauf. Dies ist auch beim R-orig zu sehen, welches beim PSNR im unendlichen Bereich liegt und sich beim NIQE auf ähnlicher Höhe mit den anderen TPs befindet, was wiederum gegen diese neue Metrik spricht.

Es kann erneut keine Beziehung zwischen beiden Metriken hergestellt werden, um dadurch die NIQE-Indexe definieren zu können.

Interessant bei dem Plot der NIQE-Indexe aller TPs ist jedoch die Erkenntnis, dass man feststellen kann, wo ein Kurzvideo aufhört und das nächste beginnt. So erzielt der Kurzfilm an der TP-Stelle 14, welches einen Sonnenaufgang hinter Bäumen und Wolken zeigt, deutlich niedrigere NIQE-Indexe im Vergleich zur gesamten Sequenz. Dies gilt ebenso für eine ähnliche Teilaufnahme der TP-Stellen von 80–94. Dies ist mithilfe der Betrachtung des PSNR-Plots nicht ganz so eindeutig zu sehen.

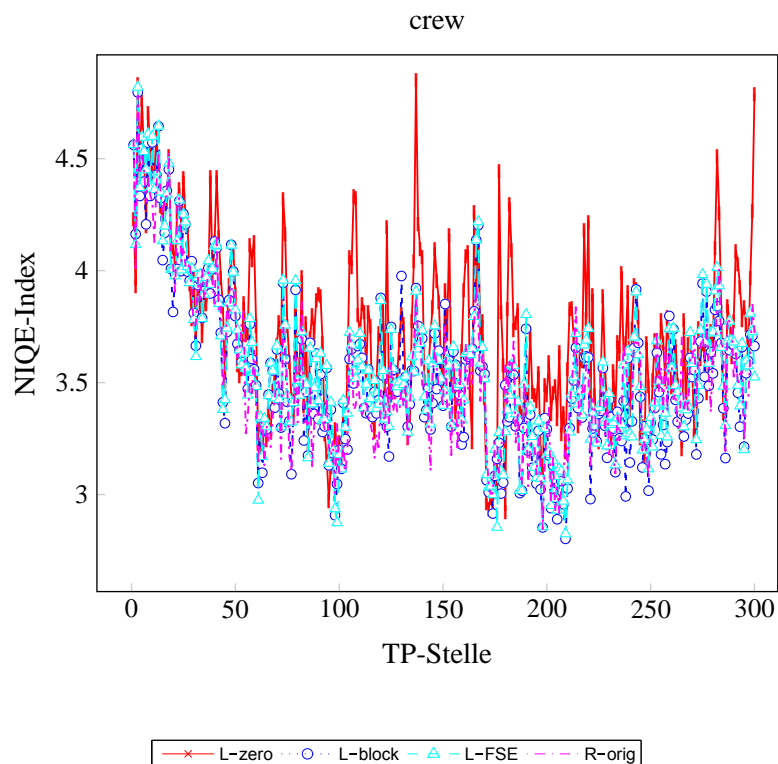


Abbildung 4.12: Ergebnisse der NIQE-Metrik von der Crew-Sequenz

4.2.3 Crew

Die wesentlichen Eigenschaften der Crew-Sequenz sind in Kapitel 4.1.3 beschrieben.

4.2.3.1 NIQE

Auf den ersten Blick fällt auf, dass die NIQE-Indexe während der gesamten Sequenz bei allen TPs im Vergleich zu den anderen Videosequenzen eher niedrig sind. Die NIQE-Indexe für L-block und L-FSE liegen fast durchgehend auf gleicher Höhe, dennoch gibt es TP-Stellen an denen sie voneinander abweichen, wenn auch nur bis zu max. 0.2. Dies ist z. B. an den TP-Stellen 7, 20, 77 und 83 zu vernehmen, siehe hierzu Abb. 4.12. In jeweils einem der Frames, aus welchen diese TPs gebildet werden, ist eine Veränderung der Helligkeit aufgrund eines Blitzlichtes zu erkennen. Das L-zero weist stets, und teilweise viele, unscharfe Stellen auf, während im L-block Artefakte vorhanden sind. Das L-FSE enthält die Artefakte ebenfalls in der gleichen Region, aber nicht so gravierend. L-block erzielt jedoch an den genannten

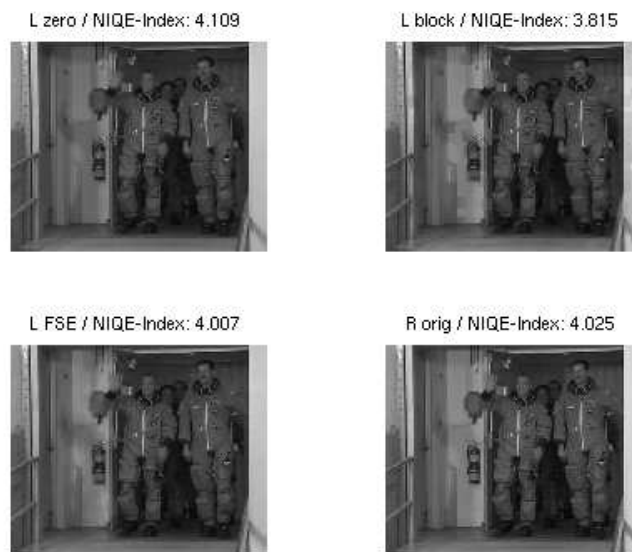


Abbildung 4.13: TP-Bilder der TP-Stelle 20 der Crew-Sequenz

TP-Stellen meist die niedrigsten NIQE-Indexe, während die Rangfolge für die anderen beiden TPs mit dem R-orig wechselt, allerdings weichen die NIQE-Index nie stärker als um 0.2 voneinander ab. Beispielsweise werden durch die Blitzlichter und die dadurch entstehenden Unterschiede bezüglich der Helligkeit Artefakte im L-block hervorgerufen, was wiederum dazu führt, dass das L-FSE nicht perfekt rekonstruiert werden kann. Demnach können Bildinformationen im L-FSE vorhanden sein, welche im R-orig nicht existieren, wie es z. B. in Abb. 4.13 links neben der Hand des vorderen, linken Crew-Mitglieds der Fall ist, verursacht durch dessen Bewegung. Demnach könnte man davon ausgehen, dass diese Abweichungen bei den NIQE-Indexen durch vorhandene Blitzlichter in einem der beiden Frames hervorgerufen werden.

Innerhalb der nächsten TP-Bilderreihe an den TP-Stellen 122, 123 und 124 steigt der NIQE-Index für das mittlere L-zero, also an der TP-Stelle 123, plötzlich stark an. Bei näherer Betrachtung aller TP-Bilder der genannten TP-Stellen ist aber keine große Veränderung zu erkennen, lediglich an der TP-Stelle 124 ist wieder ein sehr schwaches Blitzlicht enthalten. Dies korrespondiert auch mit den NIQE-Indexen, die sich, mit Ausnahme der TP-Stelle 123

für L-zero, für alle TPs nur geringfügig unterscheiden. Das L-block enthält nur ganz leichte Artefakte bei den ersten beiden TP-Bildern, beim dritten, mit Einfluss des schwachen Blitzlichtes, sind die Artefakte etwas stärker vorhanden, der NIQE-Index sinkt dagegen um 0.2 auf 3.3. Im L-FSE existieren diese Artefakte ebenso und es ist dazu noch etwas verzerrt, die NIQE-Indexe i. H. v. 3.3 jedoch bleiben für alle drei TPs gleich. Kein visueller Unterschied ist bei L-zero zu vernehmen. Dieses wirkt an allen drei TP-Stellen gleich verschwommen und dennoch erzielt es beim mittleren TP den höchsten NIQE-Index dieser TP-Bilderreihe von 4.199 und ist damit um ca. 0.7 höher als vorher und nachher. Das R-orig zeigt schwankende Ergebnisse, jedoch lediglich innerhalb eines Bereichs von max. 0.3 und konkurriert mit L-block bzw. L-FSE. Es stellt sich nun die Frage, weshalb der NIQE-Index für das mittlere L-zero erheblich ansteigt, obwohl erst an der TP-Stelle 124 ein dezentes Blitzlicht auftritt. Dieses Ergebnis steht daher im Konflikt mit der eben aufgestellten Vermutung bezüglich der untersuchten TP-Stellen 7, 20, 77, 83. Diese TP-Bilderreihe zeigt, dass erhebliche Abweichungen bei den NIQE-Indizes auch dann existieren, wenn diese Blitzlichter nur leicht bzw. gar nicht vorhanden sind.

Interessant sind die Ergebnisse an der TP-Stelle 130, welche sich aus den Frames 259, dem Referenzbild f_{2t-1} , und 260, dem aktuellen Bild f_{2t} , der Originalsequenz bilden. Sie wird deshalb untersucht, da an dieser TP-Stelle bei den NIQE-Indizes eine Abweichung von 0.4–0.6 auszumachen ist. Die beiden Frames unterscheiden sich bezüglich der Helligkeit gravierend durch ein sehr starkes Blitzlicht. L-zero ist wiederum an einigen Stellen unscharf und erzielt einen NIQE-Index von 3.814. L-block enthält extreme Artefakte aufgrund des Kontrasts beider Frames und hat an dieser TP-Stelle den höchsten NIQE-Index von 3.976. Im völlig verzerrten L-FSE sind ebenfalls viele Artefakte vorhanden, aber hier fällt der NIQE-Index von 3.413 am kleinsten aus, wofür keine Erklärung gefunden werden kann.

Ebenso herausragend zeigt sich die TP-Stelle 151. Auch hier ist ein Blitzlicht in Frame 301, dem zugehörigen Referenzbild f_{2t-1} , zu vernehmen. L-zero zeigt sich wie gehabt etwas unscharf und schneidet hierbei mit einem NIQE-Index von 3.919 am schlechtesten ab. L-block enthält wiederum sehr viele Artefakte und wird nach der NIQE-Metrik mit 3.851 bewertet. Das L-FSE ist auch hier mit 3.463 am besten, obwohl es dieselben Stellen verzerrt aufweist

sowie Artefakte enthält. Dieses Ergebnis ist erneut nicht nachvollziehbar und kann nicht erläutert werden.

Da das L-zero durchgehend unscharf abgebildet wird, scheint eine nähere Betrachtung der TP-Bilder an der TP-Stelle 164 sehr interessant. Hierbei befindet sich die Crew bereits direkt vor der Kamera, ein Blitzlicht ist nicht auszumachen und die NIQE-Indexe für L-block sowie L-FSE liegen bei fast durchschnittlichen 3.65. Das bedeutende an dieser TP-Stelle ist, dass hier das L-zero den besten NIQE-Index von 3.242 hat und das mit einem für diese Sequenz deutlichen Unterschied zu den beiden kompensierten TPs von ca. 0.3, vgl. Abb. 4.12. TP-Bilder ohne Blitzlicht wurden bereits bei den TP-Stellen 122 und 123 untersucht, jedoch kann keine Übereinstimmung zu diesem Ergebnis gefunden werden.

Die TP-Stelle 176, mit enthaltenem Blitzlicht im aktuellen Bild f_{2t} , dem Frame 352, ist aufgrund einer Differenz bei den NIQE-Indexen zwischen L-block und L-FSE bedeutend. Die Rangfolge entspricht, ungeachtet dem R-orig, der den TPs an der TP-Stelle 130. Bei näherer Betrachtung dieser TP-Bilder ist die visuelle Qualität ebenfalls dieselbe. Demnach steht auch hier das völlig verzerrte L-FSE mit starken Artefakten an erster Stelle mit einem für diese Sequenz äußerst niedrigen NIQE-Index von 2.854. Darauf folgt mit einem NIQE-Index von 3.023 das unscharf dargestellte L-zero und am schlechtesten fällt das mit Blockartefakten bedeckte L-FSE mit 3.159 aus. Die NIQE-Indexe sind deutlich kleiner als an der TP-Stelle 130, obwohl sich die Crew nun näher an der Kamera befindet. Für diese TP-Stelle hätten die NIQE-Indexe erwartungsgemäß viel höher sein müssen.

An der TP-Stelle 180 weichen die NIQE-Indexe abermals stärker voneinander ab und auch hier existiert ein Blitzlicht im Referenzbild f_{2t-1} , dem Frame 359. Nicht nachvollziehbar erzielt das stark verschwommene L-zero einen generell sehr kleinen NIQE-Index von 2.927. Als Nächstes folgt L-FSE mit 3.28 wie gewohnt mit Artefakten und verzerrt, während L-block mit erheblichen Artefakten und einem NIQE-Index von 3.489 am schlechtesten wekommt. Ein derartiger Unterschied der NIQE-Indexe zwischen L-FSE bzw. L-block und dem L-zero war bereits an der TP-Stelle 151 zu sehen, allerdings fehlt die Übereinstimmung der Rangfolge zur Interpretation für dieses Ergebnis.

Ein sehr hoher NIQE-Index für diese Sequenz wird von L-zero an der letzten TP-Stelle 300



Abbildung 4.14: TP-Bilder der TP-Stelle 300 der Crew-Sequenz

erzielt, der bei 4.793 liegt. Sie ist deshalb erwähnenswert, da das L-zero durchschnittlich um ca. 1.2 schlechter ausfällt als die beiden kompensierten TPs. Betrachtet man die letzten beiden Frames dieser Sequenz, so ist bei Frame 600, dem hierfür aktuellen Bild f_{2t} , ein Blitzlicht zu erkennen, welches im Hintergrund von Metall reflektiert wird und worauf sich zusätzlich noch ein größerer Schatten auszeichnet. Aufgrund dieser Reflexion bilden sich in dieser Region viele störende Artefakte bei L-block, dies wurde bei L-FSE nicht gut rekonstruiert, dennoch fällt L-FSE mit 3.525 geringfügig besser aus als L-block mit 3.663, siehe zur Veranschaulichung Abb. 4.14.

Eine Auswertung des besten bzw. schlechtesten TPs der gesamten Crew-Sequenz liefert keine interpretierbaren Ergebnisse zur neuen Metrik. Das gemäß dem NIQE qualitativ hochwertigste TP-Bild mit einem NIQE-Index von 2.802 ist das L-block an der TP-Stelle 209, vgl. Abb. 4.15. Da sich die Crew direkt vor der Kamera unter Einwirkung von Blitzlicht befindet und daher das L-block erhebliche Artefakte aufweist, ist dieses Ergebnis nicht nachvollziehbar. Ebenso gering fällt der NIQE-Index für L-FSE aus, der mit 2.825 nicht viel höher liegt und die visuelle Qualität des zugehörigen TP-Bildes ebenso schlecht ist. Selbst das stark

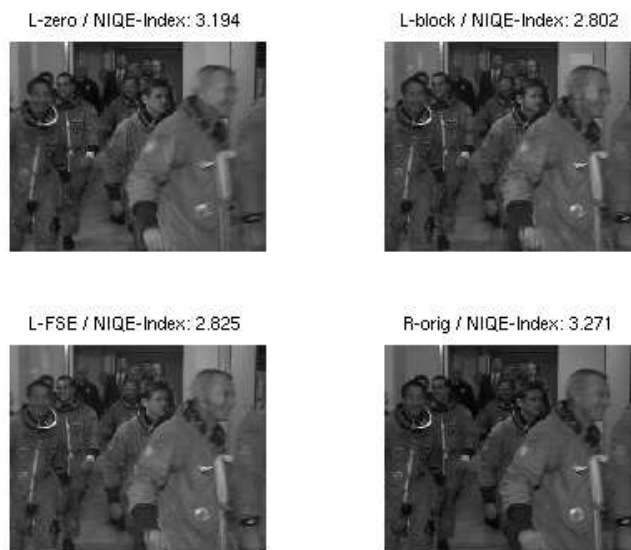


Abbildung 4.15: Bestes TP-Bild von L-block an der TP-Stelle 209 der Crew-Sequenz

unscharfe L-zero erzielt einen für diese Sequenz niedrigen NIQE-Index i. H. v. 3.194.

Betrachtet man nun das schlechteste TP-Bild an der TP-Stelle 137 von L-zero mit 4.856, so bleibt die Erklärung ebenso offen. Auch wenn das L-zero durchgehend unscharf abgebildet ist, so ist ein Unterschied zu den NIQE-Indizes von L-block bzw. L-FSE i. H. v. ca. 1 doch sehr auffallend. Die NIQE-Indexe fallen an dieser TP-Stelle generell sehr hoch aus, allerdings stellt sich hierbei die Frage nach dem Grund. Die visuelle Qualität dieser TP-Bilder scheint, aufgrund fehlender Einwirkung eines Blitzlichtes, besser zu sein und Blockartefakte sind auch nicht auszumachen.

Anhand dieser Ergebnisse lässt sich auch für diese Sequenz keine präzise Aussage zur Interpretation der neuen Metrik treffen. Außerdem kommt nun die Frage auf, ob die anhand der Basketball-Sequenz aufgestellte Vermutung, der NIQE-Index erhöhe sich, sobald sich die gefilmte Szene näher an der Kamera befindet, widerlegt werden muss. Für Aufklärung soll erneut das PSNR dienen.

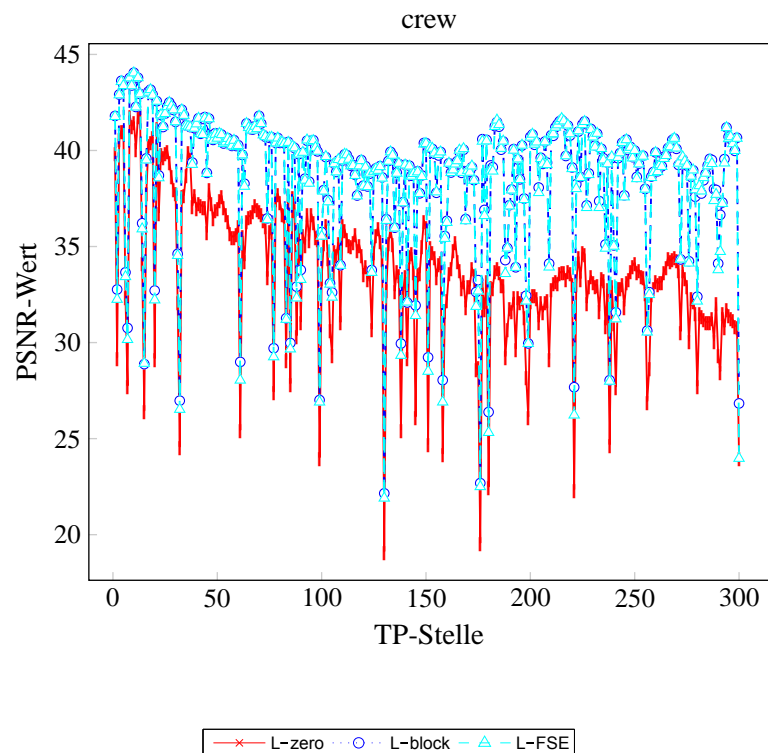


Abbildung 4.16: Ergebnisse des PSNRs von der Crew-Sequenz

4.2.3.2 PSNR

Die Auswertung der Crew-Sequenz anhand des PSNRs ergab eindeutige Ergebnisse. Die PSNR-Werte für L-block und L-FSE liegen ausnahmslos für die gesamte Sequenz auf gleicher Höhe, wodurch diese Ergebnisse gemeinsam dem L-zero gegenübergestellt werden können. Das L-zero liegt durchschnittlich um ca. 2–3 dB unterhalb der beiden kompensierten TPs an TP-Stellen, die sich aus Frames zusammensetzen, die geringes bis extremes Blitzlicht aufweisen. Für TP-Stellen, die kein Blitzlicht in einem der beiden originalen Frames enthalten, fallen die PSNR-Werte für alle TPs deutlich höher aus. Hierbei befinden sich die PSNR-Werte für L-block bzw. L-FSE bei ca. 38–40 dB, während das L-zero um ca. 5–7 dB niedriger liegt.

Die Rangfolge der TPs ist gewissermaßen dieselbe, demnach erzielt L-block nahezu durchgehend die höchsten PSNR-Werte, während sich die für das L-FSE unmittelbar darunter befinden und L-zero kontinuierlich die kleinsten PSNR-Werte erzielt. Dies wiederum kann mit

der fehlenden Berücksichtigung des PSNRs von Artefakten in Verbindung gebracht werden, die vor allem für diese Sequenz verstärkt an TP-Stellen, beeinträchtigt durch die Blitzlichter, auftreten.

Die unter Kapitel 4.2.3.1 untersuchten Extremstellen werden nun näher mit den erzielten PSNR-Werten betrachtet.

Die beim NIQE zu Beginn ausgewerteten TPs, aufgrund voneinander abweichenden Ergebnissen des L-blocks und L-FSEs, ergeben keine herausragenden PSNR-Werte. Festzuhalten ist, dass die PSNR-Werte unterdurchschnittlich ausfallen. Aufgrund existierender Blitzlichter an diesen TP-Stellen weist das L-block bei allen TP-Bildern Artefakte vor, diese treten ebenso bei L-FSE auf, welches zusätzlich etwas verzerrter, je nach Stärke der Blitzlichter, abgebildet ist. Das L-zero erzielt mit vielen unscharfen Stellen die niedrigsten PSNR-Werte. Auch harmoniert die TP-Bilderreihe an den TP-Stellen 122, 123 und 124 besser mit den PSNR-Werten als der NIQE. Wie bereits erwähnt, ist bei den ersten beiden TP-Bildern kein Blitzlicht auszumachen, jedoch tritt im Frame 248, hierbei dem aktuellen Bild f_{2t} , ein Schwaches auf. Demnach liegen die PSNR-Werte der TPs an den TP-Stellen 122 und 123 für L-block bzw. L-FSE nachvollziehbar um ca. 5 dB höher als an der TP-Stelle 124, für L-zero liegt dieser Unterschied bei ca. 3 dB.

Ebenso eindeutig sind die Ergebnisse für die TP-Stellen 130, 151, 176 und 180. Bei allen TP-Bildern ist ein sehr starkes Blitzlicht vorhanden, was zu erheblichen Blockartefakten und Verzerrungen aufgrund des Kontrasts bei L-block führt. L-FSE weist an denselben Stellen im TP-Bild Artefakte auf und ist stärker verzerrt als sonst, teilweise können auch einige Stellen nur äußerst schlecht rekonstruiert werden. Die PSNR-Werte für L-block bzw. L-FSE fallen beide gleich niedrig aus und sind um ca. 10–17 dB kleiner als durchschnittlich. Das L-zero wird umso unschärfer abgebildet, je näher sich die Crew an der Kamera befindet und ist ca. 9–14 dB schlechter als bei TP-Bildern ohne Blitzlicht.

Interessant ist auch beim PSNR die letzte TP-Stelle 300. Wie schon beim NIQE erläutert, wird beim letzten Frame 600, also dem aktuellen Bild f_{2t} , das Blitzlicht einer Kamera durch das Metall im Hintergrund reflektiert. Bemerkenswert an diesem Ergebnis ist, dass das L-FSE mit einem PSNR-Wert von 23.97 dB am schlechtesten ausfällt und L-zero mit

Crew	L-zero	L-block	L-FSE	R-orig
Mittelwert NIQE als Index	3.7246	3.5349	3.5646	3.5266
Mittelwert PSNR in dB	33.6790	38.5699	38.4859	infinity

Tabelle 4.3: Mittelwerte Crew-Sequenz

23.99 dB, wenn auch nur geringfügig besser, wekommt, vgl. hierzu Abb. 4.16. Das L-block enthält aufgrund der Reflexion des Lichtes starke Artefakte, die beim PSNR jedoch keine Berücksichtigung finden und daher den hierbei höchsten PSNR-Wert von 26.83 dB erzielt. Projiziert man das vom NIQE am besten bewertete TP-Bild auf die PSNR-Werte, so widersprechen sich die Ergebnisse. Die PSNR-Werte fallen aufgrund eines vorhanden Blitzlichtes an der TP-Stelle 209 für alle TPs schlecht aus und liegen um ca. 4–5 dB unter dem Durchschnitt. Ebenso gegensätzlich trifft dies für das gemäß dem NIQE am schlechtesten bewertete TP-Bild an der TP-Stelle 137 zu. Da in beiden Frames kein Blitzlicht existiert, fallen beim PSNR die Ergebnisse überdurchschnittlich hoch aus und liegen für L-block bzw. L-FSE bei ca. 39 dB.

Auffällig bei der Auswertung der Extremwerte des PSNRs ist, dass diese exakt mit dem Auftreten von Blitzlichtern übereinstimmen und die PSNR-Werte an diesen TP-Stellen für alle TPs rapide sinken. Demnach korrespondieren die TP-Bilder mit den PSNR-Werten miteinander. L-zero wird umso verschwommener je näher sich die Crew der Kamera zuwendet und auch bei den PSNR-Werten ist ein Rückgang zu vernehmen.

4.2.3.3 NIQE vs. PSNR

4.2.3.3.1 Betrachtung der Mittelwerte Anhand der Mittelwerte des NIQEs, einzusehen in Tab. 4.3, ist zu erkennen, dass sich die NIQE-Indexe bei der Crew-Sequenz nicht stark voneinander unterscheiden. Das R-orig liegt mit einem NIQE-Index von 3.5 auf gleicher Höhe mit L-block bzw. L-FSE und konkurriert damit stärker mit den beiden TPs als beim PSNR, wo sich das R-orig wie gewohnt im Unendlichen befindet. Das L-zero fällt auch beim NIQE am schlechtesten mit einem um ca. 0.2 höheren NIQE-Index aus. Die Auswertung des

Originalvideos mit 600 Frames ergab gegenüber dem R-orig keine Veränderung des NIQE-Indexes.

Obwohl sich die TP-Bilder der verschiedenen TPs teilweise stark unterscheiden, besonders dann, wenn ein Blitzlicht vorhanden ist, ähneln sich die NIQE-Indexe sehr. Auch ist auffällig, dass die Mittelwerte bei dieser Sequenz sehr niedrig ausfallen. Erwartungsgemäß hätten die NIQE-Indexe an den TP-Stellen mit Blitzlichtern viel höher ausfallen müssen, da hierbei der Kontrast am größten ist, welcher Einfluss auf die Varianz σ hat. Da die Varianz σ im MVG-Modell Berücksichtigung findet und damit der Unterschied zum natürlichen MVG-Modell erheblich größer wird, hätte dies auch eine sehr viel schlechtere Auswirkung auf den NIQE-Index haben müssen. Dies ist jedoch nicht der Fall ist und die Ergebnisse somit nicht verwertbar.

Die Analyse der Mittelwerte für das PSNR ist um einiges aufschlussreicher. Demnach unterscheiden sich die PSNR-Werte für L-block bzw. L-FSE kaum und befinden sich durchschnittlich um ca. 5 dB, und erwartungsgemäß, oberhalb von L-zero.

4.2.3.3.2 Betrachtung der Plots Auf den ersten Blick des Plots des PSNRs fällt auf, dass das L-zero eine Tendenz zum Schlechteren aufweist. Die PSNR-Werte werden im fortgeschrittenen Video zunehmend kleiner, was damit begründet werden kann, dass sich die gefilmte Szene immer näher der Kamera zuwendet. Dies war bereits bei der Basketball-Sequenz zu vernehmen. Bei dieser Sequenz jedoch sind die Auswirkungen gravierender, da die Blitzlichter von der Crew umso stärker reflektiert werden, je näher sich die Männer an der Kamera befinden. Bis auf das letzte TP befindet sich L-zero durchgehend unterhalb der anderen beiden TPs. Die Differenz zu den Kurven von L-block bzw. L-FSE, welche während der gesamten Sequenz übereinander liegen, wird allerdings zunehmend größer. Dass sich die Crew immer näher der Kamera zuwendet hat einen kleineren Einfluss auf das PSNR von L-block bzw. L-FSE. Eine Analyse der Extremwerte, also der immer wieder auftretenden Extremwerte zum Schlechteren, ergibt, dass es sich hierbei ausschließlich um TPs handelt, welche aus Frames mit Blitzlichtern gebildet werden. Je stärker das vorhandene Blitzlicht ist, umso niedriger fallen die PSNR-Werte für alle TPs aus.

Vergleicht man nun die Plots beider Metriken, vgl. Abb. 4.12 und Abb. 4.16, so ist beim NIQE sogar eine leichte Tendenz aller TPs zum Besseren im fortgeschrittenen Video wahrnehmbar, was sich jedoch völlig mit den korrespondierenden TP-Bildern widerspricht. Des Weiteren geht beim NIQE nicht eindeutig hervor, wann es sich um TP-Bilder mit Blitzlichtern handelt. Die NIQE-Indexe verringern sich für diese TPs nicht wesentlich, obwohl sich die einzelnen TP-Bilder doch sehr unterscheiden und die visuelle Qualität deutlich abnimmt. Die NIQE-Indexe für L-zero, welches für die gesamte Sequenz unscharf abgebildet ist, variieren stets innerhalb von ca. 2.9 bis 4.8, dennoch befindet es sich meist oberhalb der Kurven für L-block bzw. L-FSE. Wie auch bei den anderen Sequenzen zu sehen war, liegt das R-orig mit L-block bzw. L-FSE auf nahezu gleicher Höhe, schwankt jedoch stärker. Die NIQE-Indexe weichen im Grunde während der gesamten Sequenz nicht stark voneinander ab und fallen im Vergleich zu den anderen Sequenzen niedriger aus, selbst die TP-Bilder mit Blitzlichtern weisen einen nicht allzu viel höheren NIQE-Index auf als sonst.

Beim Vergleich des vom NIQE als bestes bzw. schlechtestes bewertetes TP-Bild mit dem PSNR ist das Resultat genau das Gegenteil. Demnach liegen die PSNR-Werte, beim gemäß dem NIQE qualitativ schlechtesten TP-Bild, über dem Durchschnitt. Was das als bestes Bewertetes betrifft, so liegen die PSNR-Werte für alle TPs um ca. 4 dB unter dem Mittelwert, aufgrund eines auftretenden Blitzlichtes.

Aufgrund fehlender Übereinstimmung beider Metriken bzgl. der Plots, der Kurvenverläufe sowie der Extremwerte und deren teilweise gegensätzliche Ergebnisse, wird die Eignung des NIQEs zur Qualitätsbewertung bezweifelt.

4.2.4 Flower

Die wesentlichen Eigenschaften der Flower-Sequenz sind in Kapitel 4.1.4 beschrieben.

4.2.4.1 NIQE

Diese Sequenz erfährt keine beachtlichen Veränderungen und auch die NIQE-Indexe unterscheiden sich hierbei nicht stark. Demnach liegen die NIQE-Indexe aller TPs innerhalb von

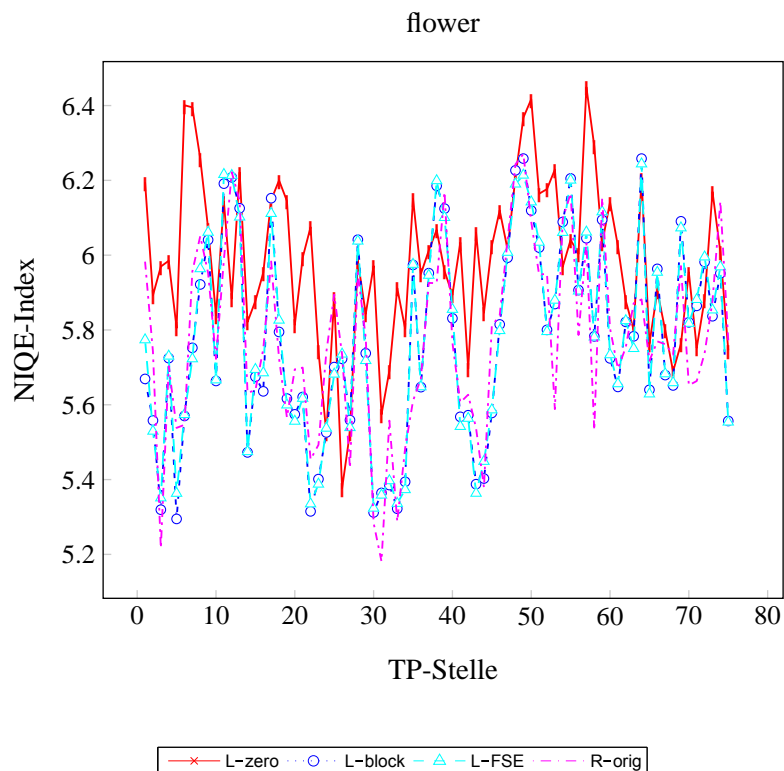


Abbildung 4.17: Ergebnisse der NIQE-Metrik von der Flower-Sequenz

5.2 und 6.4, weshalb keine herausragenden Extremwerte vorzufinden sind.

Bei diesem Video fällt allerdings auf, dass das L-zero durchgehend verschwommen ist und dies mit den NIQE-Indexen korrespondiert, welche für das L-zero nahezu kontinuierlich am höchsten sind, siehe dazu rote Kurve in Abb. 4.17. Dass bei L-zero die Umrisse des Baumes sowie das Blumenbeet immer unscharf abgebildet werden, erinnert in erster Linie an die Schwankungen der Schärfe im Vorder- bzw. Hintergrund eines TP-Bildes, wie es bereits bei der Basketball-Sequenz zu sehen war. Demnach befinden sich der Baum sowie das Blumenbeet im Vordergrund, während der Häuserblock im Hintergrund stets schärfer dargestellt wird.

Die visuelle Qualität für L-block, L-FSE sowie R-orig ist generell identisch und entsprechend an den NIQE-Indexen zu erkennen, die für die gesamte Sequenz sehr ähnlich ausfallen.

Betrachtet man das vom NIQE am besten bewertete Bild, es handelt sich hierbei um das

zugehörige Referenzbild f_{2t-1} R-orig an der TP-Stelle 31, mit dem NIQE-Index von 5.181, so ist kein visueller Unterschied zu L-block bzw. L-FSE erkennbar, was sich ebenso bei den NIQE-Indizes widerspiegelt. Lediglich das unscharfe L-zero schneidet mit 5.6 schlechter ab. Was das schlechteste TP-Bild, von L-zero, mit einem NIQE-Index von 6.4 an der TP-Stelle 57 betrifft, erzielen auch die anderen TPs im Vergleich zum Rest der Sequenz etwas höhere NIQE-Indizes. Aufgrund visueller Ähnlichkeit der Bilder für die gesamte Sequenz bleibt die Begründung hierfür aus.

Bei der Auswertung der Flower-Sequenz anhand der NIQE-Metrik fällt besonders auf, dass sich L-block, L-FSE sowie R-orig optisch gesehen nicht unterscheiden. L-block bzw. L-FSE sind beständig mit kaum wahrnehmbaren Artefakten abgebildet und beide TPs sowie das R-orig sind an übereinstimmenden Stellen, besonders im Blumenbeet im Vordergrund, unscharf. Auch sind die NIQE-Indizes für L-block bzw. L-FSE und R-orig stets nahezu identisch. Das kontinuierlich verschwommene L-zero erzielt immerzu die höchsten NIQE-Indizes.

Die neue Metrik erweckt zunächst den Eindruck, als würde sie verwertbare Ergebnisse liefern und die Unschärfe, eine für diese Sequenz überwiegende Störung, nachvollziehbar berücksichtigen. Dies soll nun mithilfe des PSNRs überprüft werden.

4.2.4.2 PSNR

Bezüglich der PSNR-Werte von L-zero schwanken die Ergebnisse eher kaum und erzielen, bis auf das Ende der Sequenz, nur leicht variierende PSNR-Werte. Lediglich für die letzten beiden TP-Stellen schneiden die PSNR-Werte schlecht ab und vergrößern daher den PSNR-Wertbereich, der für L-zero bei 21.3–24.1 dB liegt. PSNR-Werte für L-block bzw. L-FSE hingegen unterliegen stärkeren Schwankungen und befinden sich zwischen 29 und 33.5 dB. Schenkt man den Extremwerten beider Metriken Beachtung, ist festzuhalten, dass sie sich teilweise widersprechen. Sozusagen ist an der TP-Stelle 27 beim PSNR ein schlecht bewertetes TP aufzufinden, vgl. 4.18, während dieses mithilfe des NIQEs für überdurchschnittlich gut befunden wird. Die TP-Stelle 63 ist ebenfalls nicht aufschlussreich, die nach dem NIQE als eher durchschnittlich gilt und beim PSNR wiederum sehr gut ausfällt. Das gemäß dem

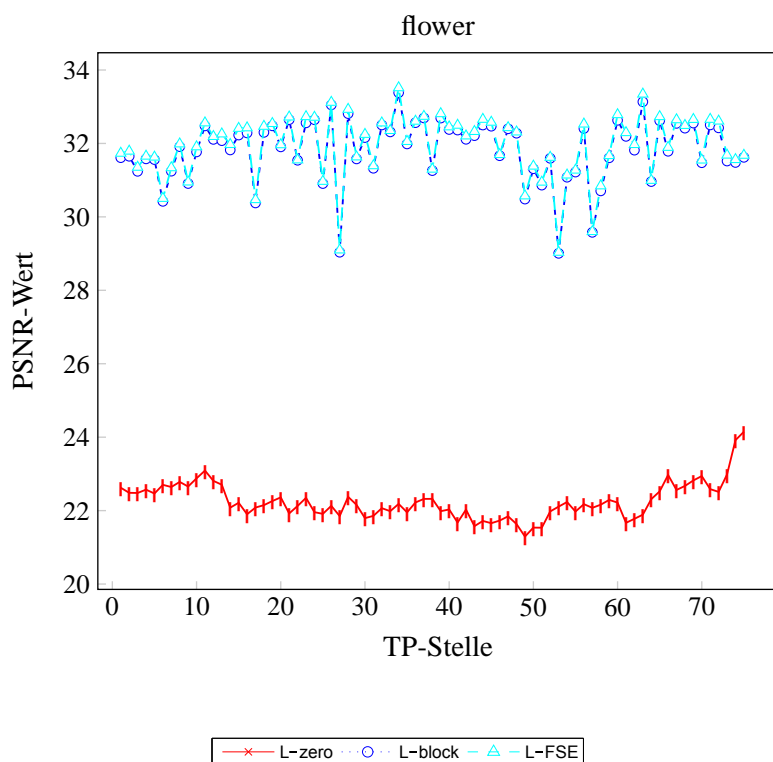


Abbildung 4.18: Ergebnisse des PSNRs von der Flower-Sequenz

Flower	L-zero	L-block	L-FSE	R-orig
Mittelwert NIQE als Index	5.9811	5.7744	5.7772	5.7773
Mittelwert PSNR in dB	22.2556	31.8338	31.9258	infinity

Tabelle 4.4: Mittelwerte Flower-Sequenz

NIQE beste TP-Bild an der TP-Stelle 31 liegt beim PSNR eher im Durchschnitt, es hebt sich also nicht vom Rest der Sequenz ab. Was das am schlechtesten bewertete TP an der TP-Stelle 57 anbelangt, so kann eine Übereinstimmung dahingehend gefunden werden, dass sich die PSNR-Werte für L-block bzw. L-FSE als drittschlechteste der Sequenz auszeichnen.

4.2.4.3 NIQE vs. PSNR

4.2.4.3.1 Betrachtung der Mittelwerte Die Mittelwerte der NIQE-Metrik haben für R-orig, L-block sowie L-FSE nahezu identische NIQE-Indexe und betragen alle, siehe dazu

Tab. 4.4, bis zur zweiten Nachkommastelle 5.77. Lediglich das L-zero grenzt sich etwas ab. Wendet man das Originalvideo mit allen 150 Frames auf die Metrik an, so erhöht sich der Mittelwert im Vergleich zu R-orig minimal auf 5.789. Da es sich bei R-orig um das originale Referenzbild f_{2t-1} handelt, ist nicht klar, weswegen es auf gleicher Höhe mit L-block bzw. L-FSE liegt.

Beim PSNR befindet sich das R-orig wie gewohnt im unendlichen Bereich. L-block bzw. L-FSE weichen erwartungsgemäß geringfügig voneinander ab und L-zero mit dem geringsten Mittelwert folgt wie gehabt zum Schluss. Ausschlaggebend für diese Rangfolge dürfte hierbei, aufgrund nahezu fehlender Artefakte bei L-block bzw. L-FSE, welche ohnehin beim PSNR keine Auswirkungen haben, die Unschärfe bei L-zero sein, weshalb es um durchschnittlich ca. 9.5 dB schlechter ausfällt.

4.2.4.3.2 Betrachtung der Plots Bei der Betrachtung des NIQEs, siehe Abb. 4.17, variieren die NIQE-Indexe, zwar innerhalb eines kleinen NIQE-Indexbereichs, aber dafür stark. Nahezu durchgehend ist ein Auf und Ab bei allen TPs zu sehen. Da aber bei näherer Betrachtung der TP-Bilder kein visueller Unterschied wahrzunehmen ist, können diese NIQE-Indexsprünge nicht erklärt werden.

Bei der Plotbetrachtung, vgl. Abb. 4.18, lässt sich sagen, dass sich das PSNR für L-zero wiederum durchgehend unterhalb der anderen TPs befindet, während L-block bzw. L-FSE einander entsprechen. Der Kurvenverlauf für L-zero zeichnet hier eher beständig aus, was durchaus mit den TP-Bildern von L-zero korrespondiert, welche stets verschwommen abgebildet und sonst keinen gravierenden Veränderungen unterlegen sind.

Die Unschärfe trifft ebenso bei L-block bzw. L-FSE zu, insbesondere in Bildregionen des Blumenbeets im Vordergrund, allerdings schwanken die PSNR-Werte für die kompensierten TPs etwas stärker. Demnach ist das Auf und Ab bezüglich L-block bzw. L-FSE beider Metriken zumindest bedingt übereinstimmend, jedoch decken sich die Ergebnisse im Hinblick auf das beste bzw. schlechteste TP-Bild abermals nicht. Auch, dass das R-orig auf gleicher Höhe mit L-block bzw. L-FSE liegt ist erneut nicht erklärbar.

Es lässt sich dadurch keine endgültige Aussage zur Interpretation der NIQE-Indexe treffen,

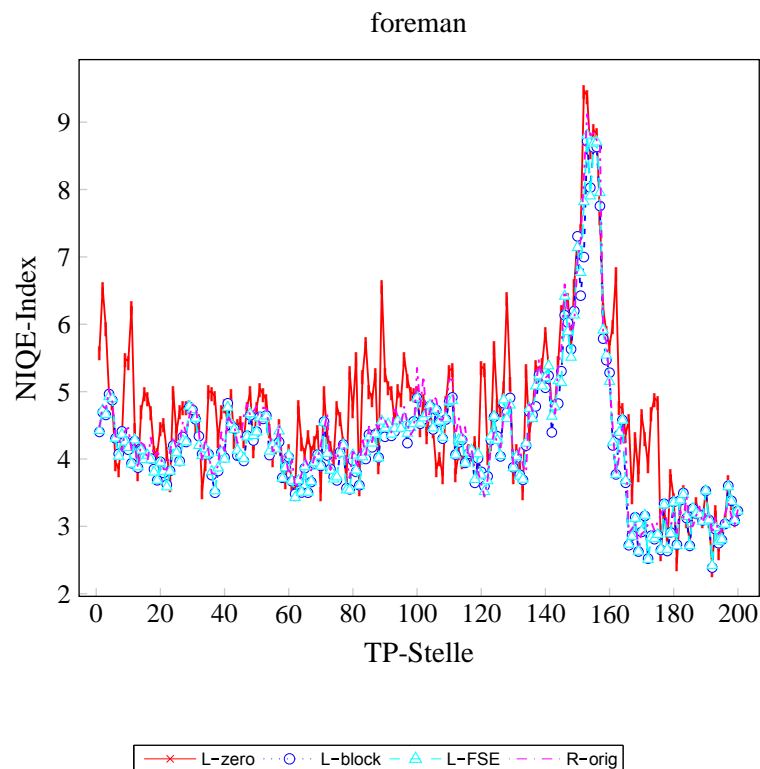


Abbildung 4.19: Ergebnisse der NIQE-Metrik von der Foreman-Sequenz

allerdings sind die Ergebnisse eher nachvollziehbar, in Anbetracht dessen, dass für diese Videosequenz hauptsächlich unscharfe Stellen, aus visueller Sicht, auszumachen sind.

4.2.5 Foreman

Die wesentlichen Eigenschaften der Foreman-Sequenz sind in Kapitel 4.1.5 beschrieben.

4.2.5.1 NIQE

Dieses Video beinhaltet lediglich kurz vor Schluss gravierende Veränderungen, wo die Kamera gedreht wird. Das ist ebenfalls sehr deutlich an den NIQE-Indexen zu erkennen, die für alle TPs ab TP-Stelle 135 erheblich ansteigen. Demnach befinden sich diese vor der Drehung für L-zero stets innerhalb eines Bereichs von 3.5 bis 6.5. Somit hat der Unterschied vom schlechtesten zum besten L-zero einen Wert von 3, was möglicherweise mit den Veränderungen der Gesichtszüge bzw. der Kopfbewegungen des Bauarbeiters, sowie dem unruhigen

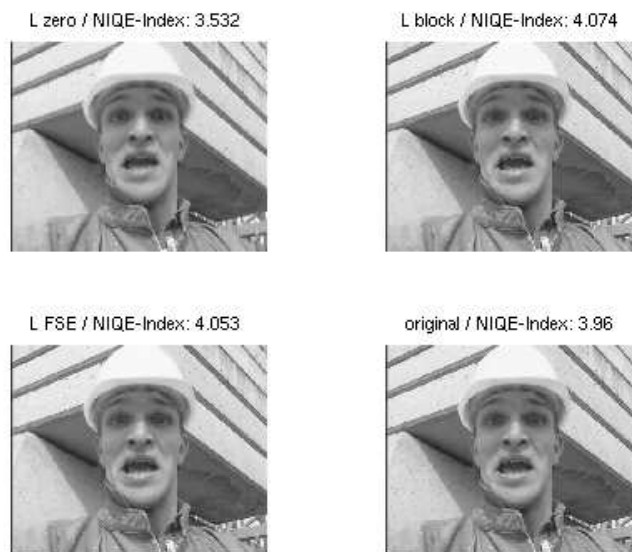


Abbildung 4.20: TP-Bilder der TP-Stelle 33 der Foreman-Sequenz

Halten der Kamera begründet werden kann. Dass das L-zero kontinuierlich leicht bis teilweise stark verschwommen ist, kann u. a. für dieses Ergebnis ausschlaggebend sein. Dagegen spielen sich die Ergebnisse für L-block bzw. L-FSE innerhalb eines NIQE-Indexbereichs von 3.4–5 ab, welche durchaus schärfer abgebildet sind und dennoch ähneln sich die NIQE-Indexbereiche von L-zero und den kompensierten TPs sehr. Aus diesem Grund werden, in Bezug auf die Untersuchung herausragender NIQE-Indexe, bis zur Kameradrehung zwei TP-Stellen näher betrachtet, an denen das L-zero den kleinsten NIQE-Index erzielt, was deutlich in Abb. 4.19 zu sehen ist.

Zum einen handelt es sich um die TP-Stelle 33. Die Unschärfe bei L-zero ist ganz deutlich am Gesicht zu erkennen, die Steinblöcke im Hintergrund unterscheiden sich allerdings nicht zu den anderen TP-Bildern. Trotzdem wird es hierbei als bestes TP-Bild, vgl. Abb. 4.20 mit einem NIQE-Index von 3.532 gewertet und liegt im Durchschnitt um ca. 0.5 niedriger. Zum anderen stechen die Ergebnisse der Bilderreihe an den TP-Stellen 106–108 hervor, an denen L-zero jeweils mit durchschnittlich 3.87 erneut als bestes TP abschneidet. Es zeigen sich zwar deutlich weniger unscharfe Stellen und dennoch ist der NIQE-Index nun um ca. 0.7

niedriger im Vergleich zu den anderen TPs, die optisch gesehen besser abgebildet sind. Der NIQE-Index fällt trotzdem, wenngleich besserer Optik, höher als an der TP-Stelle 33 aus. Da an diesen TP-Stellen wiederum hauptsächlich unscharfe Stellen auszumachen sind, kommt hierbei die Frage auf, inwieweit das nun beste Abschneiden des verschwommenen L-zero mit den Ergebnissen der Flower-Sequenz in Verbindung gebracht werden kann.

Nun zur Betrachtung der Ergebnisse, welche sich aus der Kameradrehung ergeben. Hier ist eine deutliche Verschlechterung der Qualität aller TPs an den TP-Stellen 136–160 auszumachen. Bei diesen TPs sind alle TP-Bilder verschwommen und L-zero, wessen NIQE-Indexe hier sogar bis zu ca. 9.5 reichen, ist noch unschärfer als sonst. Die NIQE-Indexe für L-zero liegen an diesen TP-Stellen stets leicht oberhalb von L-block bzw. L-FSE. Sobald die Kamera wieder ruhig auf den Rohbau gerichtet wird, verringern sich auch die NIQE-Indexe für alle TPs wieder. Ab der TP-Stelle 176 liegen die NIQE-Indexe für L-zero, L-block und L-FSE auf gleicher Höhe und aufgrund der ruhigen Kamerahaltung ist auch hier das L-zero schärfer abgebildet.

Die Analyse des vom NIQE am besten bewertete TP-Bild an der TP-Stelle 192 kann aus visueller Sicht gut nachvollzogen werden. Die Kamera wird hierbei ruhig auf den Rohbau gerichtet und die NIQE-Indexe fallen für diese TP-Bilder, bei denen kein Unterscheid wahrzunehmen ist, ähnlich niedrig aus. Der kleinste NIQE-Index i. H. v. 2.371 wird von L-zero erzielt, dicht gefolgt von L-block bzw. L-FSE. Allerdings ergibt sich für das R-orig die schlechteste Bildqualität mit einem NIQE-Index von 2.888 und liegt damit um ca. 0.5 höher als das L-zero. Eine mögliche Erklärung hierfür könnte sein, dass durch die Mittelung beim L-zero zwischen zwei originalen Frames, bei denen sich das Bild, vom Rauschen abgesehen, nicht wesentlich ändert, der Rauschanteil verkleinert wird. Demnach enthält im Vergleich das korrespondierende Referenzbild mehr Rauschen, was ein Grund für das schlechte Ergebnis von R-orig sein könnte.

Das TP-Bild mit der schlechtesten Qualität befindet sich erwartungsgemäß im Bereich der Kameradrehung. Der höchste NIQE-Index wird hierbei erneut von L-zero mit 9.471 erreicht und obwohl die anderen TP-Bilder ebenfalls sehr unscharf abgebildet sind, liegen diese NIQE-Indexe zwischen 6.9 und 8.2. Allerdings ist bei L-zero am meisten Unschärfe

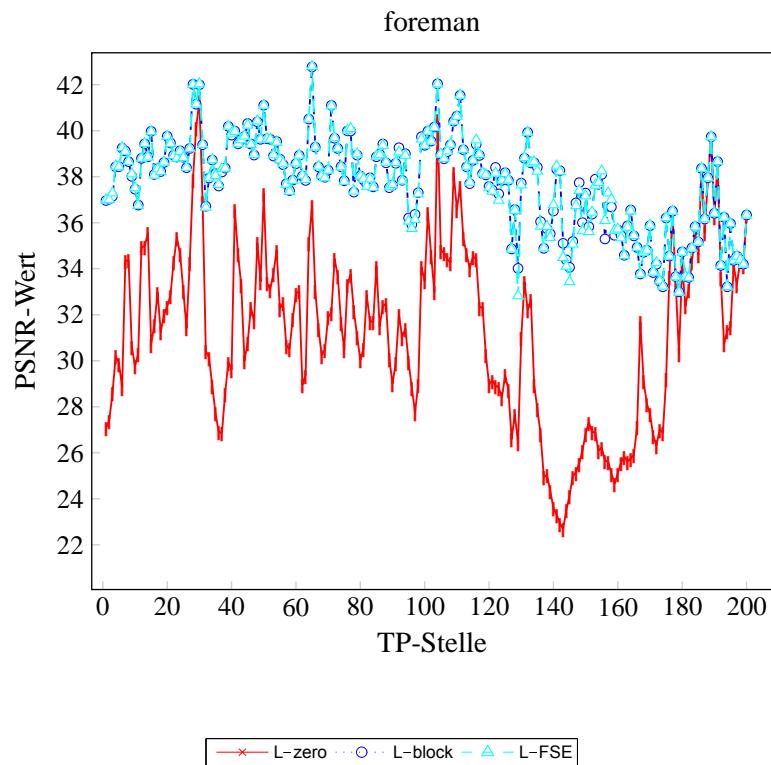


Abbildung 4.21: Ergebnisse des PSNRs von der Foreman-Sequenz

erkennbar, z. B. wird der Kran im Hintergrund äußerst schlecht dargestellt.

4.2.5.2 PSNR

Zur Hilfestellung der Ergebnisanalyse der Foreman-Sequenz wird wieder das PSNR herangezogen und bei näherer Betrachtung ist erkennbar, dass die TP-Bilder besser mit den PSNR-Werten korrespondieren. Hierbei befindet sich das stets leicht bis teilweise stark unscharfe L-zero durchgehend um durchschnittlich ca. 6 dB unterhalb der anderen beiden TPs. Das Ende des Videos kann hiervon ausgenommen werden, wo die Kamera in ruhiger Haltung den Rohbau filmt, denn gemäß den PSNR-Werten liegt das L-zero ab diesem Bereich auf gleicher Höhe mit L-block und L-FSE, vgl. rote Kurve in Abb. 4.21.

Für die gesamte Sequenz erzielen L-block bzw. L-FSE mit nahezu identischen PSNR-Werten die besten Ergebnisse, während das L-zero starken Schwankungen unterlegen ist. Die extremen Unterschiede bei L-zero können dahingehend begründet werden, dass diese TP-Bilder

manchmal optisch gut, aber auch sehr oft nur völlig unscharf abgebildet werden. Je nachdem wie viele Bewegungen des Bauarbeiters sowie Positionsveränderungen der Kamerahaltung vorhanden sind, beeinflusst dies die Schärfe des L-zeros negativ. Das R-orig liegt wie gewohnt im unendlichen Bereich.

Auch bei dieser Qualitätsmetrik ist an den PSNR-Werten zu erkennen, dass sich diese bei der Kameradrehung ab der TP-Stelle 135 für alle TPs, insbesondere L-zero, deutlich verringern. Sobald die Kamera gegen Ende des Videos wieder ruhig gehalten wird, verändern sich auch die PSNR-Werte entsprechend.

Das gemäß dem NIQE für gut befundene L-zero an der TP-Stelle 33, welches viele unscharfe Stellen aufweist, erzielt nach dem PSNR ein anderes Ergebnis. Demnach wird die Unschärfe dieses TP-Bildes für die Qualitätsbeurteilung berücksichtigt und dementsprechend niedrig fällt der PSNR-Wert aus. Diese unscharfen Stellen sind beim L-block bzw. L-FSE nicht aufzufinden, daher liegen die PSNR-Werte hierbei im Durchschnitt.

Bei Untersuchung der Bilderreihe an den TP-Stellen 106–108 konnten keine Übereinstimmungen gefunden werden. Beim NIQE fällt die Reihe auf, da das L-zero an allen drei TP-Stellen den niedrigsten NIQE-Index erzielt. Was jedoch das PSNR anbelangt, so wird dieser Beurteilung nicht entsprochen. Hier liegen die PSNR-Werte für alle TPs über dem Durchschnitt mit der erwarteten Rangfolge, wonach L-FSE dicht gefolgt von L-block höhere PSNR-Werte erreichen und L-zero jeweils zwischen ca. 4–5 dB darunter liegt.

Die Überprüfung des besten NIQE-Indexes dieser Sequenz ergibt, dass sich die PSNR-Werte für das beste TP von L-zero nicht ebenso ergeben. An der TP-Stelle 192 zeigt die Kamera in ruhiger Position auf den Rohbau und schneidet mithilfe des PSNRs gut ab und liegt für alle TPs über dem Durchschnitt. Jedoch handelt es sich nach dem PSNR um kein herausragendes Resultat.

Das gemäß dem NIQE qualitativ schlechteste TP-Bild an der TP-Stelle 152 von L-zero, wo die Kamera die extreme Drehung erfährt, wird ebenfalls schlecht nach dem PSNR bewertet, demnach liegt dieses mit ca. 4.3 dB unterhalb der durchschnittlichen Ergebnisse für das L-zero. Allerdings fallen die PSNR-Werte innerhalb dieser Kameradrehung für alle TPs eher niedrig aus, ein auffallend schlechter PSNR-Wert ist hierbei nicht aufzufinden.

Foreman	L-zero	L-block	L-FSE	R-orig
Mittelwert NIQE als Index	4.6698	4.2178	4.2345	4.3544
Mittelwert PSNR in dB	31.2044	37.7280	37.7154	infinity

Tabelle 4.5: Mittelwerte Foreman-Sequenz

Dennoch ist eine kleine wahrnehmbare Korrespondenz beider Metriken zu vernehmen, was das beste bzw. schlechteste TP anbelangt, wenn auch nur bedingt. Demnach findet der NIQE sowohl als auch das PSNR das schlechteste TP-Bild im Bereich der Kameradrehung und das beste TP-Bild gegen Ende der Sequenz, wo die Kamera wieder ruhig gehalten wird.

4.2.5.3 NIQE vs. PSNR

4.2.5.3.1 Betrachtung der Mittelwerte Werden die Mittelwerte der verschiedenen TPs ausgewertet, so erhält man beim NIQE für L-block und L-FSE nahezu identische NIQE-Indexe i. H. v. 4.2, gefolgt von R-orig mit 4.3544 und zu guter Letzt das L-zero. Lediglich das L-zero mit einem ca. 0.4 höheren Mittelwert grenzt sich etwas ab, wie der Tab. 4.5 zu entnehmen ist, was vermutlich auch hier mit den teilweise sehr unscharfen TP-Bildern zu begründen ist. Der Mittelwert des NIQEs, angewandt auf das Originalvideo mit 400 Frames, ergibt keine Veränderung in Bezug auf den von R-orig.

Beim PSNR stehen L-block sowie L-FSE erneut mit ähnlichen PSNR-Werten von 37.7 dB an erster Stelle und L-zero schneidet am schlechtesten ab. Diese Rangfolge entspricht der Erwartung.

4.2.5.3.2 Betrachtung der Plots Der Plot für L-zero bei der NIQE-Metrik unterliegt sehr großen Schwankungen während der gesamten Sequenz und es ist ein kontinuierliches Auf und Ab erkennbar, was mit der sich stets variierenden Schärfe des L-zeros begründet werden kann. Dennoch befindet sich die rote Kurve von L-zero im Grunde oberhalb von L-block bzw. L-FSE. Bei der Kameradrehung werden alle TP-Bilder zunehmend unschärfer und die NIQE-Indexe für alle TPs entsprechend schlechter. Zum Schluss liegt L-zero auf annähernd gleicher Höhe mit den Kurven für L-block und L-FSE, welche wiederum zusammen aus-

gewertet werden können, da diese annähernd gleichauf liegen und denselben Kurvenverlauf vorweisen.

Im Vergleich zum PSNR liegt L-zero durchgehend unterhalb den anderen Kurven und unterliegt auch hier großen Schwankungen. Die Begründung hierfür ist die gleiche wie für NIQE. Ebenso verringern sich auch hier die PSNR-Werte für alle TPs bei der Kameradrehung und zum Schluss ist wiederum ein leichter Anstieg dieser erkennbar.

Wenn auch keine gemeinsamen Extremwerte auffindbar sind, so ist durchaus ein ähnlicher Kurvenverlauf der TPs beider Metriken auszumachen, siehe Abb. 4.19 und Abb. 4.21.

Die Kurven für R-orig befinden sich beim PSNR im unendlichen Bereich, während diese beim NIQE mit den anderen TPs konkurriert, was wiederum gegen die Metrik spricht, zumal es sich beim R-orig um die originalen Referenzbilder f_{2t-1} handelt.

4.2.6 Discovery-Orient

Die wesentlichen Eigenschaften der Discovery-Orient-Sequenz sind in Kapitel 4.1.6 beschrieben.

4.2.6.1 NIQE

Diese Sequenz ist ähnlich der Discovery-City aufgebaut und beinhaltet unterschiedliche Kurzaufnahmen. Abermals können die TPs untereinander für das gesamte Video nicht verglichen werden, daher erfolgt eine Untersuchung innerhalb des jeweiligen Kurzfilms.

Beim ersten Video handelt es sich um einen Beduinen in der Wüste mit Sonnenstrahlen im Hintergrund, welche die Auslöser der Blockartefakte bei L-block und L-FSE sind. Solange Sonnenstrahlen direkt zur Kameralinse durchdringen, treten in dieser Region verstärkt Artefakte auf. Zoomt man bei diesen TP-Bildern den Himmel heran, so ist statt einem glatten Hintergrund leichtes Rauschen erkennbar. Das L-zero ist bei dieser Teilsequenz leicht verschwommen, da sich die Kamera langsam dreht, jedoch wird das Tier im Hintergrund schärfer abgebildet, wodurch erneut eine Beziehung zur Schärfe der TPs im Vorder- bzw. Hintergrund hergestellt werden kann. Die Auswertung dieser TPs an der TP-Stelle 8 ist aufgrund

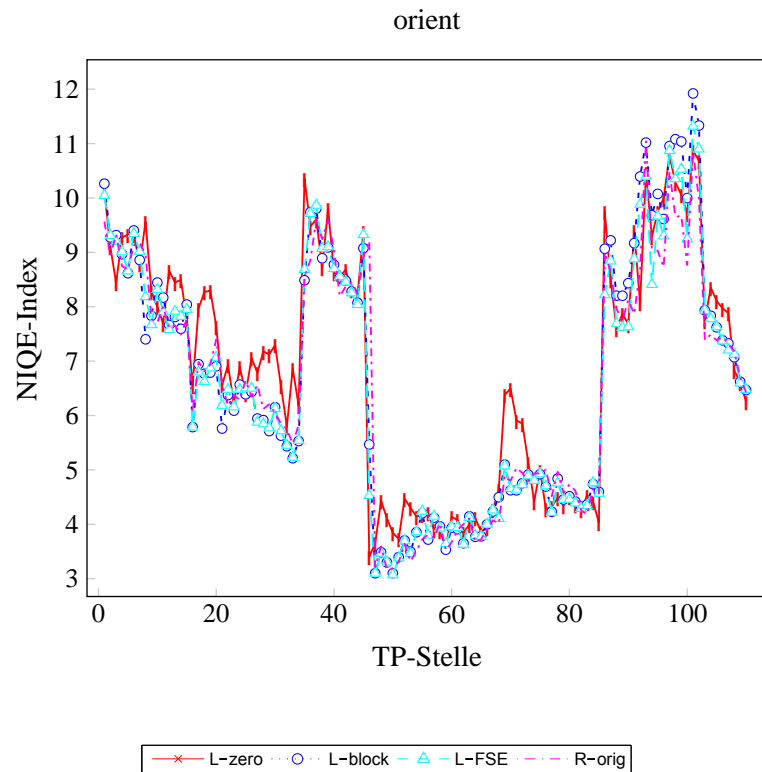


Abbildung 4.22: Ergebnisse der NIQE-Metrik von der Discovery-Orient-Sequenz

der voneinander stark abweichenden NIQE-Indexe interessant. So steht L-block mit 7.401 und einigen Artefakten an erster Stelle. Als Nächstes folgt das L-FSE, weniger Artefakte beinhaltend und etwas verzerrt, mit einem NIQE-Index von 8.185. Mit 8.95 kommt als drittes das R-orig und den Schluss mit 9.561 bildet das an dieser TP-Stelle unschärfste TP-Bild von L-zero. Die aufgestellte These hinsichtlich der Einflussnahme des Rauschens auf den NIQE muss angezweifelt werden. Am Himmel der originalen Referenzbilder f_{2t-1} R-origs ist eindeutig Rauschen wahrzunehmen. Durch die Mittelung bei L-zero nimmt das Rauschen ab, da dies beim R-orig aber bestehen bleibt, müsste dies also, ähnlich der Foreman-Sequenz, schlechter ausfallen. Dies ist hier nicht der Fall und das R-orig kommt sogar um ca. 1.5 besser weg als das L-zero.

Gleich zu Beginn des dritten Kurzvideos, an der TP-Stelle 35, zeichnet sich erneut eine TP-Stelle aus, bei welcher die NIQE-Indexe plötzlich rapide ansteigen. In dieser Kurzaufnahme wird ein kleiner Junge gezeigt, der lediglich seinen Kopf bewegt. Für dieses Video liegen

die NIQE-Indexe generell unerklärbar hoch, aber beim ersten TP sind die NIQE-Indexe am höchsten. Wichtig hierbei zu erwähnen ist, dass die TPs an dieser TP-Stelle nicht aus Frames unterschiedlicher Kurzaufnahmen berechnet werden. Bei näherer Betrachtung der TP-Bilder sind kaum Unterschiede für L-block, L-FSE sowie R-orig erkennbar. Das L-zero ist, aufgrund der Bewegung des Kindes, sehr verschwommen und erzielt nachvollziehbar den schlechtesten NIQE-Index i. H. v. 10.35. Dieser liegt um durchschnittlich 2 höher als bei den anderen TPs, bei denen die NIQE-Indexe ziemlich gleich ausfallen.

Die TP-Stelle 46 besteht aus den Frames 91 und 92, welche unterschiedlichen Kurzaufnahmen angehören. Das Referenzbild f_{2t-1} stellt das Letzte der Teilsequenz mit dem kleinen Jungen dar, während im Frame 92, dem aktuellen Bild f_{2t} , Steinkonstruktionen zu sehen sind und den Beginn des darauffolgenden Kurzvideos bildet. Innerhalb dieser folgenden Teilaufnahme sind keine Bewegungen der Bildobjekte vorhanden, lediglich die Kamera verändert die Position, wobei zum Schluss noch ein sitzender Mönch zu sehen ist. Dieses TP wurde ausgewählt, da sich die NIQE-Indexe für dieses Kurzvideo plötzlich wieder gravierend dem Besseren zuwenden, darüber beinhaltet dieses Teilvideo den besten NIQE-Index von 3.071 von L-FSE an der TP-Stelle 50. Bei den zu untersuchenden TP-Bildern an der TP-Stelle 46 handelt es sich um eine Transition, wie es schon bei der Discovery-City-Sequenz zu sehen war. Die NIQE-Indexe aller TPs weichen sehr voneinander ab. Demnach stellt das optisch beste TP-Bild L-zero dar, welches beide Frames 91 und 92 deutlich wiedergibt und scharf rekonstruiert wurde. Übergänge bei L-zero wurden bei der Discovery-City-Sequenz stets als schlechtestes TP bewertet, an dieser TP-Stelle erzielt es plötzlich mit 3.406 den niedrigsten NIQE-Index. Das L-block enthält aufgrund des Kontrasts beider Kurzfilme erhebliche Artefakte, dennoch liegt es mit 5.468 an dritter Stelle. Dazwischen befindet sich L-FSE mit 4.866, trotz Artefakten und geringerer Verzerrung. Was das R-orig anbelangt, so fällt das Ergebnis bedeutend hoch aus, zumal es sich um das originale Referenzbild f_{2t-1} , also um Frame 91, handelt und der NIQE-Index beträgt 9.213. Die Erklärung hierfür ist, dass dieses Frame noch dem vorhergehenden Kurzvideo angehört, wessen NIQE-Indexe höher liegen. Herausragend ist u. a. ein NIQE-Index von L-zero, welcher sich an der TP-Stelle 70 ergibt und um durchschnittlich 2 höher ausfällt als bei den anderen TPs, deren NIQE-Indexe na-

hezu gleichauf liegen. Bei dieser Teilsequenz sind Menschen in der Natur zu sehen, welche neben Elefanten herlaufen oder auf ihnen reiten. Es sind nur geringfügige Bewegungen auszumachen. Selbst für L-zero ist bei näherem Hinsehen kein Unterschied aufzufinden und dennoch schneidet es auffallend schlechter ab, wie deutlich aus der Abb. 4.22 hervorgeht. Abermals fehlt die Begründung der sich stark differenzierenden NIQE-Indexe, trotz nahezu gleichbleibender Bilder.

Sehr interessant sind die NIQE-Indexe an den TP-Stellen 86–102, welche sich aus dem vorletzten Video ergeben. Es zeigt im Hintergrund ein orientalisches Gebäude und wird von einer Position aus gefilmt, welche überdacht ist, sodass der Vordergrund dunkel bis schwarz abgebildet ist. Zoomt man auch hier wiederum den Himmel näher heran, ist ein Rauschen in dieser Region erkennbar. Hierbei kann nur bedingt mit der Verringerung des Rauschanteils aufgrund der Mittelung bei L-zero argumentiert werden, zumal an diesen TP-Stellen das L-zero des Öfteren oberhalb der magenta Kurve für R-orig liegt. Im Prinzip bewegt sich in dieser Teilsequenz lediglich die Kamera. Optisch gesehen sind die TP-Bilder an den jeweiligen TP-Stellen identisch, auch sind keine Artefakte bei L-block oder Rekonstruktionsfehler mit Artefakten bei L-FSE auszumachen. Dennoch schwanken die NIQE-Indexe an diesen TP-Stellen für alle TPs teilweise stark und befinden sich im höheren Bereich von 7.596–11.92. In diesem Teilvideo befindet sich zudem das TP mit dem höchsten NIQE-Index der gesamten Sequenz i. H. v. 11.92 von L-FSE an der TP-Stelle 101. Weiter auffällig war, dass das L-FSE des Öfteren an diesen TP-Stellen am schlechtesten wegkommt. Dass die NIQE-Indexe für diese Teilsequenz im höheren Bereich liegen, könnte an dem Rauschen liegen, auch was das gravierende Auf und Ab anbelangt.

4.2.6.2 PSNR

Während des gesamten Videos befinden sich die PSNR-Werte für L-zero stets um ca. 6–7 dB unterhalb L-block bzw. L-FSE, deren Ergebnisse sich nahezu durchgehend decken.

Wertet man diese Videosequenz mithilfe des PSNRs aus, so sind ebenfalls einige Extremwerte aufzufinden. Beim Vergleich dieser TP-Stellen wird schnell klar, dass es sich hierbei oft um dieselben handelt, allerdings mit teilweise umgekehrtem Ergebnis. Die TP-Stelle 8

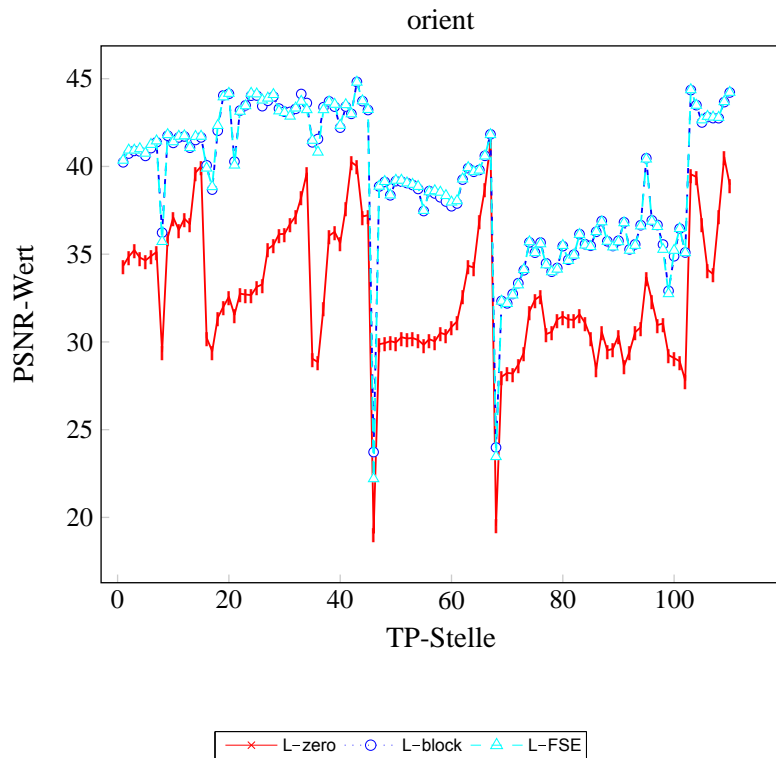


Abbildung 4.23: Ergebnisse des PSNRs von der Discovery-Orient-Sequenz

erzielt nach dem PSNR das schlechteste Ergebnis des ersten Videos, jedoch für alle TPs. Hingegen fällt beim NIQE lediglich das L-zero schlecht aus und die anderen TPs gehen an dieser TP-Stelle als das beste Ergebnis für diese Teilsequenz hervor. Im dritten Video fällt die TP-Stelle 35 gemäß dem NIQE besonders schlecht aus. Zum einen handelt es sich um den höchsten NIQE-Index für diesen Kurzfilm und zum anderen sticht er beim NIQE hervor, da das vorhergehende Kurzvideo sehr viel besser abschneidet. Was das PSNR anbelangt, so fällt lediglich der PSNR-Wert für L-zero niedrig aus und L-block sowie L-FSE liegen im Durchschnitt. Besonders auffällig ist die fehlende Übereinstimmung an der TP-Stelle 46, deren Ergebnisse anhand der NIQE-Bewertung überdurchschnittlich gut ausfallen und beim PSNR sogar das schlechteste Ergebnis der gesamten Sequenz darstellt, wie in Abb. 4.23 gut erkennbar ist. Die TP-Stelle 70 ist lediglich in Anbetracht des L-zeros von Bedeutung, dessen NIQE-Index im Gegensatz zu den anderen TPs an dieser TP-Stelle viel höher ausfällt, bei PSNR wiederum ist kein herausragendes Ergebnis auszumachen. Abermals können auch

Discovery-Orient	L-zero	L-block	L-FSE	R-orig
Mittelwert NIQE als Index	6.905	6.693	6.605	6.61
Mittelwert PSNR in dB	32.85	39.22	39.22	infinity

Tabelle 4.6: Mittelwerte Discovery-Orient-Sequenz

die Schwankungen beim NIQE der vorletzten Videoaufnahme nicht interpretiert werden, da sich die PSNR-Werte in diesem durchschnittlich und beständig verhalten.

Ein übereinstimmendes Resultat bezüglich des gemäß dem NIQE besten bzw. schlechtesten TPs besteht auch bei dieser Sequenz nicht. Des Weiteren passen zwar erneut die Extremstellen zusammen, aber das Ergebnis widerspricht sich bei dieser Sequenz vollkommen und lässt erneut keine Aussage zur neuen Metrik zu.

4.2.6.3 NIQE vs. PSNR

4.2.6.3.1 Betrachtung der Mittelwerte Sofern man zur Betrachtung der Mittelwerte das R-orig außer Acht lässt, entspricht die Rangfolge gemäß dem NIQE dem zur Qualitätsbewertung erwarteten Ergebnis, vgl. Tab. 4.6. Demnach erzielt L-FSE die niedrigsten NIQE-Indexe, dicht gefolgt von L-block und zum Schluss um ca. 0.3 höher das L-zero. Die Mittelwerte weichen für die Discovery-Orient-Sequenz nicht stark voneinander ab, jedoch muss nochmals hervorgebracht werden, dass es sich um verschiedene Teilsequenzen handelt und die NIQE-Indexe pro Kurzvideo anders ausfallen. Das R-orig befindet sich zwischen L-FSE und L-block. Der Mittelwert für die Originalsequenz, mit 220 Frames, bringt keine nennenswerte Veränderung mit sich.

Stellt man diese Resultate den Mittelwerten des PSNRs gegenüber, sie liegen L-FSE und L-block auf exakt gleicher Höhe mit einem PSNR-Wert von 39.22 dB. L-zero befindet sich um ca. 6 dB darunter und R-orig wie erwartet im Unendlichen. An den hohen PSNR-Werten ist gut erkennbar, dass es sich um eine qualitativ gute Videosequenz handelt, dies kann wiederum anhand der Mittelwerte beim NIQE nicht behauptet werden, wenn man diese mit den Mittelwerten der Crew-Sequenz, sofern dies überhaupt möglich ist, vergleicht.

4.2.6.3.2 Betrachtung der Plots In Abb. 4.22 und Abb. 4.23 sind auf den ersten Blick einige übereinstimmende Stellen vorhanden. Bei L-zero beim NIQE fällt auf, dass sich dessen Kurvenverlauf in Vergleich zu den anderen TPs mit wenigen Ausnahmen nur leicht oberhalb deren Kurven befindet, was auch den Mittelwerten, welche sich nicht stark differenzieren, zu entnehmen ist. Beim PSNR liegt L-zero durchgehend und deutlich wahrnehmbar unterhalb L-block sowie L-FSE und R-orig im Unendlichen, während dieses beim NIQE wiederum mit den anderen TPs mehr konkurriert.

In Bezug auf L-block und L-FSE, deren Kurven gleichsam übereinander liegen und aus diesem Grund gemeinsam betrachtet werden können, verhalten sich die Verläufe der kompensierten TPs beim NIQE ähnlich dem L-zero. Dieser Aussage kann ebenso beim PSNR entsprochen werden. Auffallend beim PSNR ist, dass sich die PSNR-Werte des L-blocks sowie L-FSEs an den Extremstellen dem L-zero sehr nahe kommen. Dass sich diese Extremstellen beim PSNR auch beim NIQE herauskristallisieren, wurde bereits erwähnt, wenn auch mit fehlender Übereinstimmung der Tendenz.

Wie auch der Discovery-City-Sequenz zu entnehmen ist, so kann man abermals anhand der Plots feststellen, an welcher TP-Stelle eine Kurzaufnahme aufhört und die nächste beginnt. Im Unterschied zur Discovery-City-Sequenz ist diese Eigenschaft auch besser mithilfe des PSNRs zu sehen. Diese Erkenntnis ist jedoch in Anbetracht der Interpretation des NIQE-Indexes bzw. dessen Eignung zur Qualitätsbewertung wenig hilfreich. Die Auswertung der Extremstellen spricht erneut gegen den NIQE.

4.2.7 Head

Gegen Empfehlung der Erfinder wird nun die Metrik aus der Veröffentlichung bewusst auf CT-Daten angewandt, um etwaige Unterschiede bei den Ergebnissen deuten zu können. Bei den hierfür herangezogenen CT-Daten der Head-Sequenz entspricht der maximale Pixelwert 4096, im Gegensatz zu dem bei natürlichen Bildern i. H. v. 255. Außerdem differenzieren sich die Pixelmaße, also die Werte für Width und Height. Während diese bei den CT-Daten bei jeweils 512 Pixel liegen, entsprechen sie bei natürlichen Bildern 352×288 . Da sich u. a.

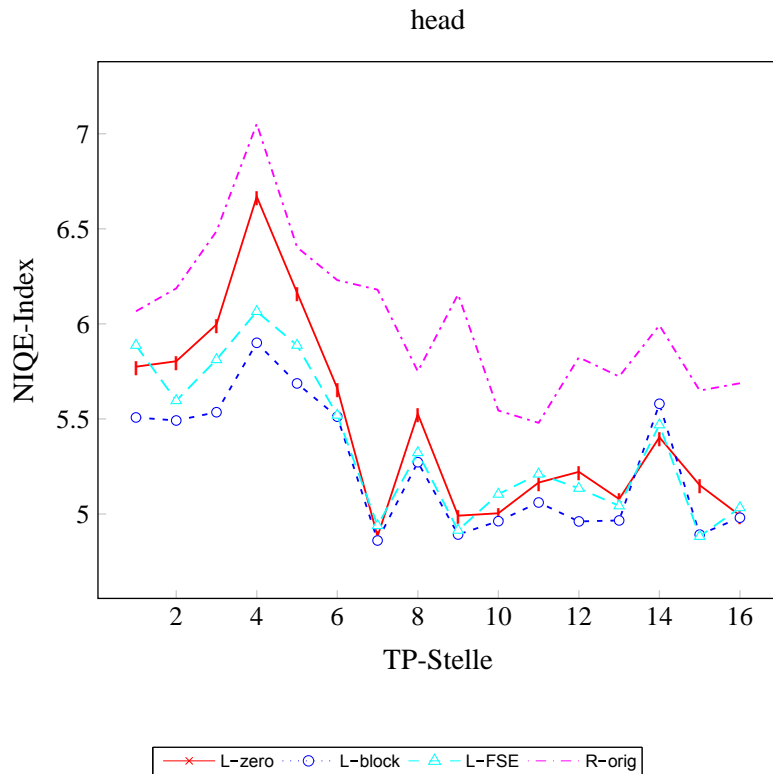


Abbildung 4.24: Ergebnisse der NIQE-Metrik von der Head-Sequenz

diese Eigenschaften von natürlichen Bildern und CT-Daten unterscheiden, soll die Auswertung die These, dass das NIQE-Modell zur Qualitätsbeurteilung für unnatürliche Bilder nicht geeignet ist, untermauern. Es handelt sich dabei lediglich um die zweite Schicht der Head-Sequenz mit insgesamt 32 Frames die nun auf die neue Metrik angewandt werden soll.

4.2.7.1 NIQE

Auf den ersten Blick sind keine Unterschiede zu den Ergebnissen beim NIQE, angewandt auf natürliche Bilder, zu vernehmen, siehe Abb. 4.24. Gleich bei der ersten TP-Stelle, dem Referenzbild f_{2t-1} , also Frame 1 der gesamten Schicht, ist eine Zahnfüllung zu erkennen. Auf den NIQE-Index scheint dies an der Stelle jedoch keine Auswirkung zu haben, denn die NIQE-Indexe verhalten sich für die gesamte Schicht ähnlich. An der TP-Stelle 4 fallen die NIQE-Indexe generell eher schlecht aus, bei Betrachtung der Bilder sind allerdings keine Besonderheiten auszumachen.

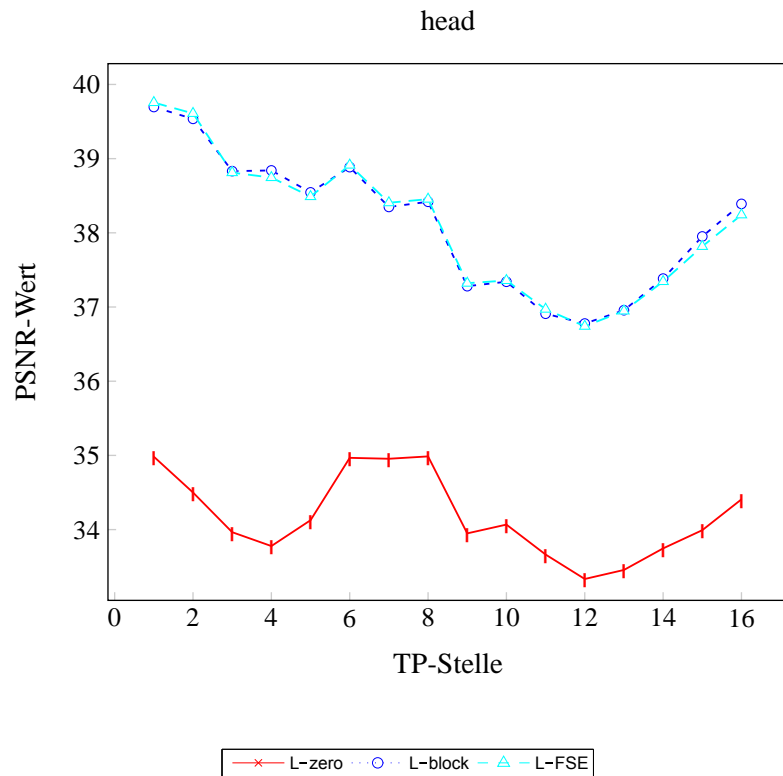


Abbildung 4.25: Ergebnisse des PSNRs von der Head-Sequenz

Wie bereits bei den Videosequenzen zu erkennen war, steht auch hier meistens das L-block an erster Stelle, während sich L-FSE und L-zero abwechseln. Auffällig ist, dass das R-orig an allen 16 TP-Stellen die höchsten NIQE-Indexe erzielt und durchschnittlich um ca. 0.66 höher ausfällt. Der Unterschied ändert sich auch nicht für die Auswertung des Originalvideos dieser Schicht der Head-Sequenz mit allen 32 Frames.

4.2.7.2 PSNR

Die Zahnfüllung im ersten Frame dieser Schicht beeinflusst das PSNR kaum. Die erste TP-Stelle beim PSNR erzielt sogar für alle TPs die höchsten PSNR-Werte der gesamten Schicht. Die Übereinstimmung bezüglich des ausschreitenden Ergebnisses an der TP-Stelle 4 beim NIQE ist bei Betrachtung des PSNRs nicht gegeben.

Während sich L-block bzw. L-FSE zwischen 36.5 und 40 dB befindet, liegt das L-zero durchgehend um ca. 3–5 dB unterhalb der kompensierten TPs. Das R-orig liegt wie gewohnt im

Head	L-zero	L-block	L-FSE	R-orig
Mittelwert NIQE als Index	5.468	5.254	5.364	6.026
Mittelwert PSNR in dB	34.18	38.13	38.12	infinity

Tabelle 4.7: Mittelwerte Head-Sequenz

unendlichen Bereich. Die Kurven, vgl. Abb. 4.25, für L-FSE bzw. L-block sind stets nahezu auf gleicher Höhe. Dieses Ergebnis entspricht der Erwartung und kann eindeutig interpretiert werden.

4.2.7.3 NIQE vs. PSNR

4.2.7.3.1 Betrachtung der Mittelwerte Die Betrachtung der Mittelwerte in Tab. 4.7 verdeutlicht nochmals die Rangfolge, die mit der beim PSNR übereinstimmt, das R-orig ausgenommen. An den hohen Werten für das PSNR ist ersichtlich, dass es sich um eine gute Qualität dieser CT-Daten handelt, während die NIQE-Indexe diesbezüglich keine Aussage zulassen und wenig aufschlussreich sind.

4.2.7.3.2 Betrachtung der Plots Vergleicht man die Plots beider Metriken, so ist für alle 16 TP-Stellen keine Übereinstimmung der Ergebnisse zu finden, siehe dazu Abb. 4.24 und Abb. 4.25. Während sich beim NIQE das L-zero geringfügig oberhalb bzw. teilweise auf gleicher Höhe mit der cyanen und blauen Kurve für jeweils L-FSE und L-block befindet, ergeben sich die PSNR-Werte hierzu schlüssig. Das L-zero liegt beim PSNR stets um ca. 4 dB unterhalb von L-block bzw. L-FSE. Alle Kurven sind beim NIQE stärkeren Schwankungen unterlegen als mit der Auswertung des PSNRs und generell ist kein kurvenähnlicher Verlauf beider Metriken erkennbar. Überdies ist am NIQE-Indexbereich nicht zu sehen, dass die Metrik hierbei auf unnatürliche Bilder angewandt wurde, dieser befindet sich für die zweite Schicht der Head-Sequenz bei 4.861–7.052.

Die Anwendung von CT-Daten auf die neue Metrik liefert keine unterscheidbaren Ergebnisse im Vergleich zu den Videosequenzen, demnach entspricht die Betrachtung der sich hierfür ergebenden NIQE-Indexe nicht der Erwartung. Es wurde auf eindeutig höhere NIQE-Indexe

spekuliert, zumal es sich bei CT-Daten nicht um natürliche Bilder handelt und aus diesem Grund nicht für die neue Metrik bestimmt sind. Dennoch weicht das Ergebnis bezüglich der Schwankungen bei den Kurven, der Rangfolge und auch beim NIQE-Indexbereich im Verhältnis zu den Videosequenzen nicht ab. Des Weiteren fehlt erneut die Übereinstimmung mit dem PSNR.

Der Behauptung, der NIQE sei nur auf natürliche Bilder anwendbar und dessen Eignung fehle für Unnatürliche, kann hiermit nicht zugestimmt werden. Die Ergebnisse sind nicht verwertbar und für die Unterstützung dieser These nicht ausreichend.

4.3 NIQE mit CT-Datenbank-Modell

Hier werden die Ergebnisse der NIQE-Metrik mit der Modellberechnung aus CT-Daten, wie in Kapitel 2.3.2 beschrieben, diskutiert. Für diese Art von Untersuchung werden verschiedene medizinische CT-Volumendaten herangezogen.

Zum besseren Verständnis soll zunächst die Begriffsverwendung erläutert werden. Bei der Sequenz handelt es sich um alle vorhandenen Schichten einer bestimmten CT-Aufnahme. Diese Sequenzen beinhalten unterschiedlich viele und unterschiedlich große Schichten. Die Größe der einzelnen Schichten bestimmt sich nach der Anzahl vorhandener Frames.

Was das spezifische Training der Metriken anbelangt, so erfolgt dies für jedes dritte Frame aller vorhandenen Schichten der jeweiligen Sequenz, beginnend beim ersten Frame.

Beim Testen der Metriken wird, beim Vergleich der Originale, für alle Schichten ebenso jedes dritte Frame zum Auswerten herangezogen, allerdings beim zweiten Frame begonnen, sodass die Metriken nicht auf exakt identische Frames trainiert und getestet werden. Hingegen werden, beim Vergleich der Originale für bestimmte Schichten, alle Frames der jeweiligen Schicht zum Testen verwendet.

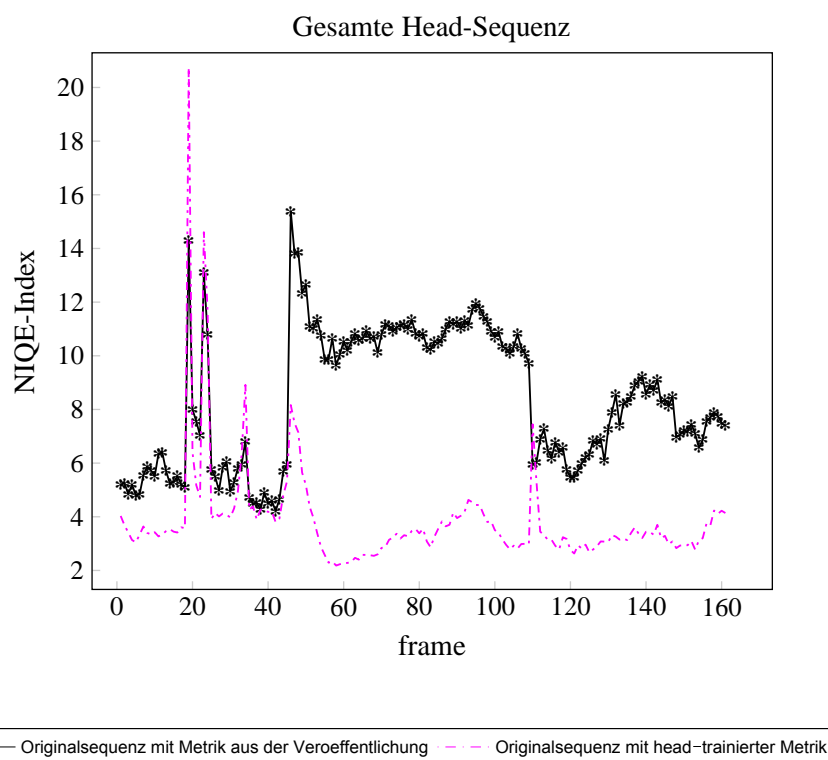


Abbildung 4.26: Ergebnisvergleich Head-Sequenz

4.3.1 Head-Trainierter NIQE vs. NIQE aus der Veröffentlichung

4.3.1.1 Vergleich Original für alle Schichten

Wie an der magenta Kurve aus Abb. 4.26 zu erkennen ist, werden nun deutlich bessere Ergebnisse für alle 5 Schichten erzielt. Ausnahmen bilden nur wenige Frames, so z. B. das Frame 19, wonach der NIQE-Index 20.8 beträgt und um ca. 7.4 höher also vorher ausfällt. Eine Untersuchung des vorhergehenden und nachfolgenden Frames ergibt jedoch keinerlei Hinweise auf Veränderungen, da es sich nahezu um das gleiche Bild handelt. Vernachlässigt man diese Extremstelle, so lässt sich aussagen, dass sich das Original vorher innerhalb eines NIQE-Indexbereichs von 4.2 bis 15.4 befand, während sich dieser nun bei 2.2 bis 14.61 bewegt. Ebenso fragwürdige NIQE-Indexe mit fehlender Erklärung ergeben sich bei den Frames 23, 24, 33, 34, 110, mit denen das Original mithilfe der head-trainierten Metrik eindeutig höhere NIQE-Indexe erzielt. Trotzdem tritt eine wesentliche Verbesserung ein. So ergibt sich beispielsweise nun ein kleinerer Mittelwert i. H. v. 3.8, welcher vorher 8.4 betrug. Überdies

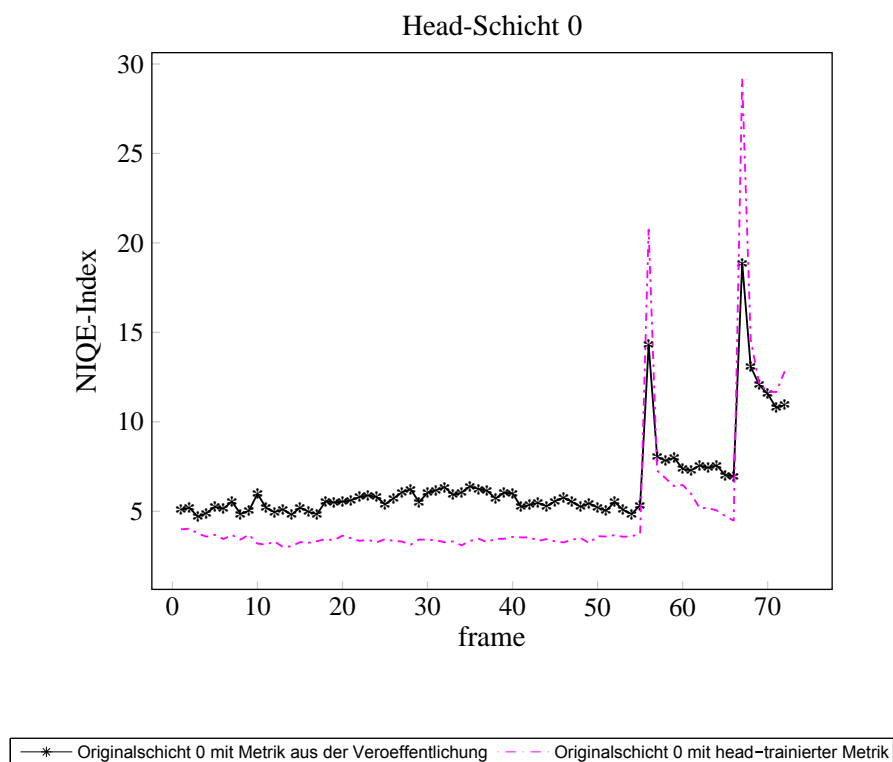


Abbildung 4.27: Ergebnisvergleich Head-Schicht 0

verhalten sich beide Kurven ähnlich.

4.3.1.2 Vergleich Original für Schicht 0

Nun zu den Ergebnissen, die sich speziell aus der 0ten Schicht der Head-Sequenz ergeben. Extremwerte sind, wie in Abb. 4.27 aufgezeigt, nur an zwei Stellen vorhanden. Beispielsweise findet sich beim Frame 56 ein um 6.44 höherer NIQE-Index für die head-trainierte Metrik als vorher. Ein näherer Hinblick auf die vorhergehenden und nachfolgenden Bilder ergibt, dass sich diese auch hier ähneln. Es handelt sich demnach um ein nahezu gleiches Bild, aber die Metrik aus der Veröffentlichung bewertet das Frame 56 mit einem um ca. 9 höheren NIQE-Index als zum vorhergehenden Bild, während sich der NIQE-Index mithilfe der head-trainierten Metrik sogar um ca. 17 erhöht. Des Weiteren sind ab dem 67ten Frame plötzlich alle NIQE-Indexe anhand der head-trainierten Metrik schlechter. Da jede Schicht mit ein und demselben medizinischen Gerät aufgenommen wurde, kann demnach auch nicht

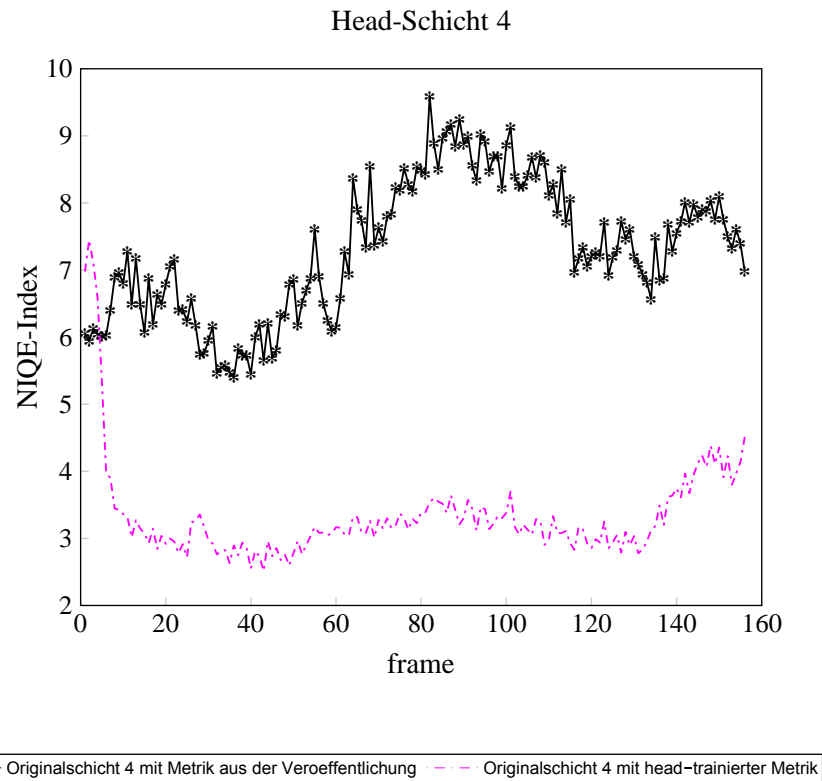


Abbildung 4.28: Ergebnisvergleich Head-Schicht 4

mit unterschiedlicher Strahlungsbelastung argumentiert werden.

Die Verbesserung wird jedoch beim NIQE-Indexbereich offensichtlich, dieser begann vorher bei 4.7 wohingegen dieser nun bei 3 liegt und bei 4.031 endet, die NIQE-Indexe ab dem 56ten Frame außen vor gelassen, um lediglich den Bereich mit beständigen NIQE-Indexen zu betrachten. Außerdem verkleinert sich der Mittelwert aller Frames von 6.5 auf 5 und beide Kurven verlaufen gleich.

4.3.1.3 Vergleich Original für Schicht 4

Betrachtet man die Bilder im Anfangsbereich, vgl. Abb. 4.28, näher, so fallen deutliche Strahlungen im Mundbereich auf. Eine Zahnfüllung ist vermutlich ausschlaggebend für diese nicht verwertbaren Ergebnisse zu Beginn der 4ten Schicht und liefert anhand der head-trainierten Metrik höhere NIQE-Indexe als vorher. Lässt man nun die ersten sechs Frames außer Acht ist erneut eine erhebliche Verbesserung der NIQE-Indexe auszumachen. Der NIQE-

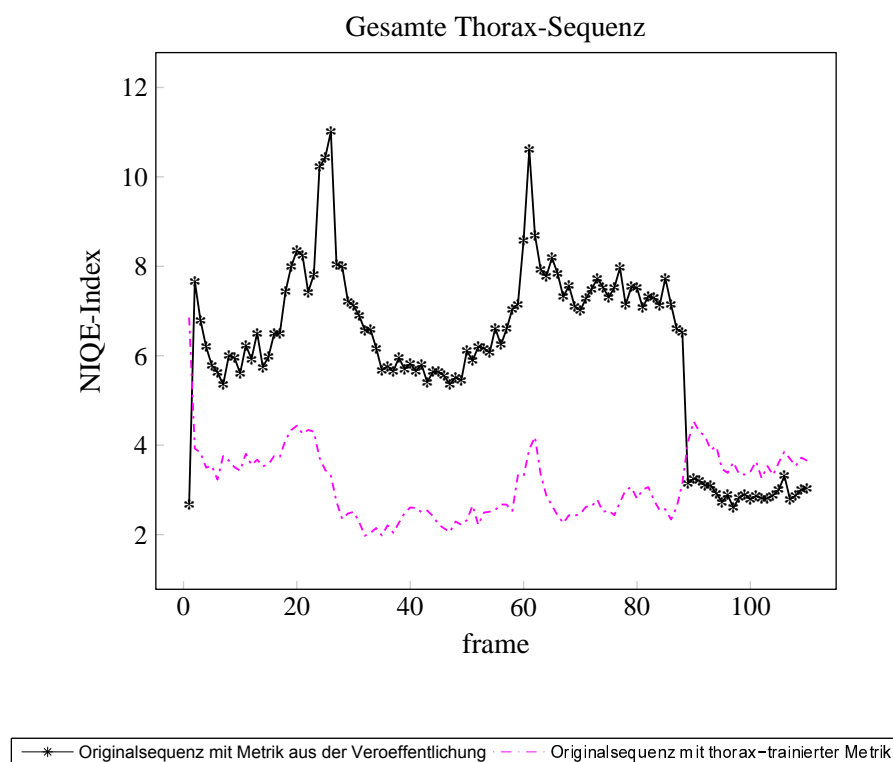


Abbildung 4.29: Ergebnisvergleich Thorax-Sequenz

Indexbereich befand sich vorher bei 5.4 bis 9.6 und liegt nun bei 2.5 bis 4.5, demnach sind auch die Schwankungen geringer geworden. Der neue NIQE-Indexbereich ähnelt stark dem für die 0te Schicht. Ein Vergleich der Mittelwerte, den Anfangsbereich wieder mit einbezogen, untermauert die Verbesserung. Der Mittelwert für die Metrik aus der Veröffentlichung liegt bei 7.3 und verringert sich erheblich auf 3.3.

4.3.2 Thorax-Trainierter NIQE vs. NIQE aus der Veröffentlichung

4.3.2.1 Vergleich Original für alle Schichten

Die erste von 5 Schichten der Thorax-Sequenz besteht lediglich aus einem Frame und die letzte Schicht beginnt ab Frame 89, was hier sehr gut in Abb. 4.29 zu sehen ist, nach der aus der 0ten und der 4ten Schicht dieser Sequenz mit der thorax-trainierten Metrik höhere NIQE-Indexe als vorher resultieren. Für die Schichten 1, 2 und 3 hingegen verkleinern sich die NIQE-Indexe und liegen hierbei stets um ca. 2 bis ca. 7 unterhalb der mit der Metrik aus



Abbildung 4.30: Ergebnisvergleich Thorax-Schicht 1

der Veröffentlichung erzielten Ergebnisse. Auch die Kurven beider Metriken verhalten sich ähnlich. Für die Schichten 1–3 ergibt sich nun ein Indexbereich von 1.974 bis 4.437, welcher vorher bei 2.6 bis 11 lag. Auch der Mittelwert, errechnet aus allen Schichten, verringert sich von ca. 6.1 auf 3.1. Wären da nicht die Ergebnisse für die erste und letzte Schicht, hätten sich für diese Sequenz schlüssige NIQE-Indexe ergeben, so bleibt zunächst eine Erklärung offen.

4.3.2.2 Vergleich Original für Schicht 1

Dieser Plot liefert ein sehr eindeutiges Ergebnis. Das Original befindet sich mit der thorax-trainierten Metrik durchgehend unterhalb der schwarzen Kurve, vgl. Abb. 4.30. Der NIQE-Indexbereich liegt nun zwischen 3.1 und 4.6. Damit sind die NIQE-Indexe stets um durchschnittlich ca. 3 geringer als vorher und erzielen einen Mittelwert von ca. 3.8. Dieses Ergebnis entspricht halbwegs den Erwartungen.

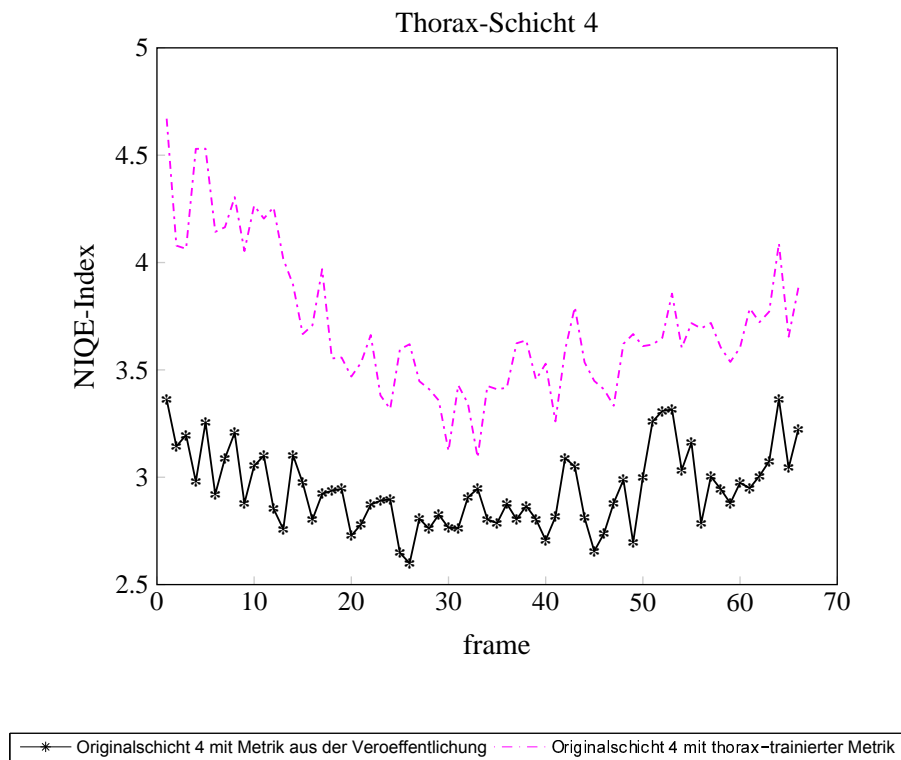


Abbildung 4.31: Ergebnisvergleich Thorax-Schicht 4

4.3.2.3 Vergleich Original für Schicht 4

Die Untersuchung der unter Abb. 4.31 aufgezeigten Kurven beider Metriken steht nun völlig im Konflikt mit dem Ergebnis aus Kapitel 4.3.2.2. Die NIQE-Indexe der 4ten Schicht der thorax-trainierten Metrik erzielen konstant um ca. 0.8 höhere Werte als vorher und das innerhalb eines NIQE-Indexbereichs von 3.1 und 4.7. Allerdings fiel mit Beobachtung in Kapitel 4.3.2.1 bereits auf, dass die NIQE-Indexe für die letzte Schicht stets höher liegen. Eine Erläuterung hierzu gibt es jedoch nicht.

Es soll nicht unerwähnt bleiben, dass sich für die 1te und die 4te Schicht die NIQE-Indexbereiche gleichen, wie es bereits bei der head-trainierten Metrik zu sehen war.

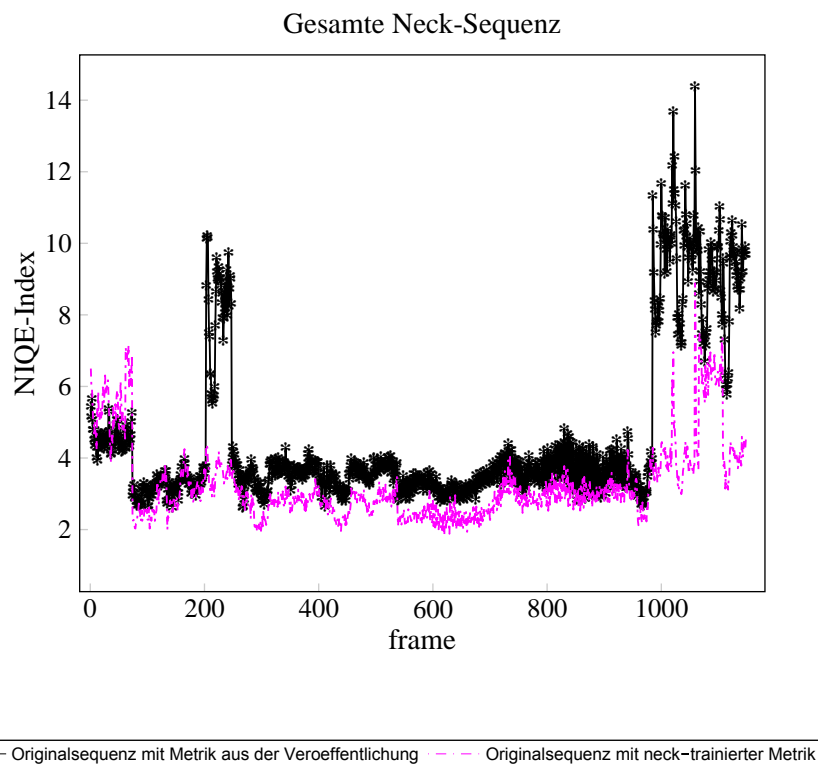


Abbildung 4.32: Ergebnisvergleich Neck-Sequenz

4.3.3 Neck-Trainierter NIQE vs. NIQE aus der Veröffentlichung

4.3.3.1 Vergleich Original für alle Schichten

Die Neck-Sequenz besteht insgesamt aus 16 Schichten, mit jeweils unterschiedlich vielen Frames. Die Abb. 4.32 zeigt ebenso eine Verbesserung der neck-trainierten Metrik, lediglich der Anfangsbereich erreicht höhere NIQE-Indexe als vorher, zu welchem die ersten 4 Schichten zählen. Betrachtet man die Bilder, so sind in Schicht 1 Frame 40 und Schicht 2 Frames 34 und 37 Zahnfüllungen zu erkennen, die Schichten 3 und 4 beginnen jedoch erst mit dem Unterkiefer. Daher kann nicht mit unverwertbaren Ergebnissen aufgrund diesen Zahnfüllungen argumentiert werden, zumal die restlichen Frames keine Besonderheiten aufweisen. Darüber hinaus sind beispielsweise in den Schichten 9 Frame 58, 10 Frame 55, 13 Frame 1, 14 Frames 1 und 2 ebenso Zahnfüllungen vorhanden, deren NIQE-Indexe befinden sich aber unterhalb der schwarzen Kurve, vgl. erneut Abb. 4.32. Besonders auffällig sind die Ergebnisse für die 8te Schicht, wo sich die NIQE-Indexe sogar um ca. 2–6 verbessern. In-

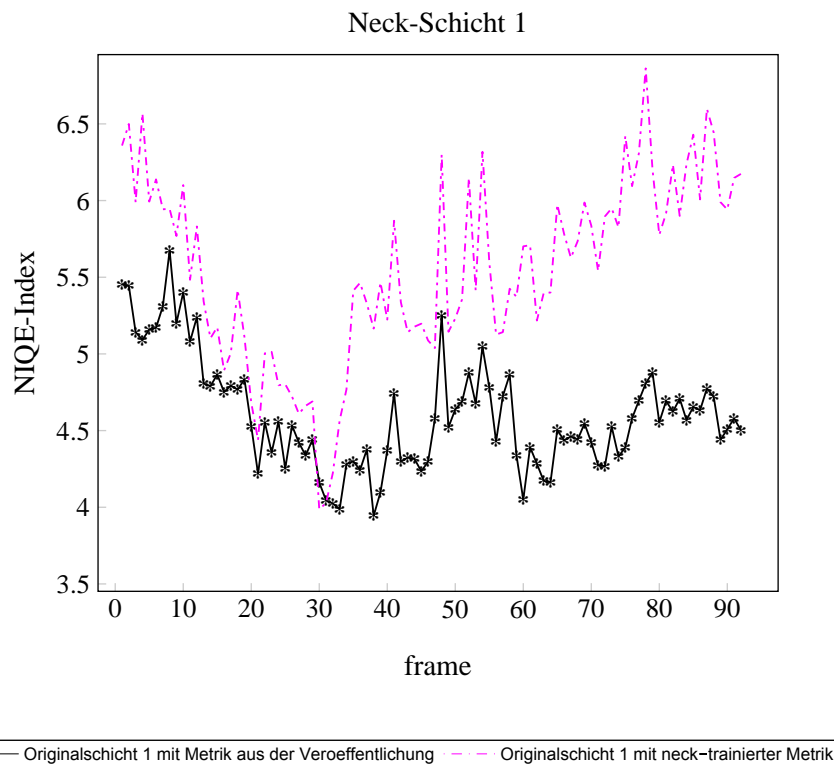


Abbildung 4.33: Ergebnisvergleich Neck-Schicht 1

nerhalb der gesamten 8ten Schicht konnten jedoch keine Eigenarten festgestellt werden, um die extrem niedrigeren NIQE-Indexe begründen zu können, welche die Ergebnisse mit der Metrik aus der Veröffentlichung derart negativ beeinflusst hätten.

Für andere Schichten wiederum verbessern sich die NIQE-Indexe durch die neck-trainierte Metrik kaum. Demnach enthält diese Sequenz sehr unterschiedliche Ergebnisse, zum einen verringern sich die NIQE-Indexe gravierend, zum anderen kaum und im Anfangsbereich sind sie höher als zuvor.

4.3.3.2 Vergleich Original für Schicht 1

Nun soll nochmal der Anfangsbereich näher betrachtet werden. Lediglich an 2 Stellen fällt die magenta Kurve unter die Schwarze, siehe hierzu Abb. 4.33. In den hierzu korrespondierenden Frames 30 und 31, ist jedoch keine Besonderheit aufzufinden.

Für diese Schicht kann allerdings vermutet werden, dass der Grund für die höheren NIQE-

Indexe anhand der neck-trainierten Metrik an der Anzahl der Frames liegt. Insofern wäre es von Bedeutung wie viele Frames die zu untersuchende Schicht beinhaltet und wie diese Anzahl im Verhältnis zu der der anderen Schichten steht. Die durchschnittliche Anzahl der Frames pro Schicht liegt bei 214 Frames, wobei die erste Schicht mit 91 Frames weit darunter liegt. Die meisten Schichten enthalten ca. 130 Frames, nur die 9te, 10te und 11te Schicht mit jeweils 433, 433 und 1210 Frames überwiegen demnach bei der Mittelung. Das würde bedeuten, dass die Metrik auf eine größere Anzahl von Frames der Schichten 9–11 trainiert wird und deshalb eine Auswertung der 1ten Schicht mit ihren 91 Frames zu keinen interpretierbaren Ergebnissen führt, da die Gewichtung dieser Frames fehlt. Ein tiefergehender Ansatz könnte sein, dass die trainierte Metrik sensibler auf bestimmte Eigenschaften dieser Schichten reagiert, z. B., auf differenzierte Strahlungsbelastungen der CT-Daten, die mit unterschiedlichen medizinischen Geräten aufgenommen wurden.

4.3.3.3 Vergleich Original für Schicht 10

Bis auf wenige Stellen liegen für diese Schicht die NIQE-Indexe wieder unterhalb der Ergebnisse nach der Metrik aus der Veröffentlichung, vgl. Abb. 4.34. Es bestehen sowohl vorher als auch nachher große Schwankungen, allerdings ist der Kurvenverlauf für schwarz und magenta gleich. Der NIQE-Indexbereich verbessert sich von 2.6–4.2 auf 1.96–3.507 und auch der Mittelwert sinkt um ca. 0.8 auf 2.725. Damit fehlt jedoch die Übereinstimmung der NIQE-Indexbereiche der näher untersuchten Schichten 1 und 10 wie es bei den anderen beiden Sequenzen festzustellen war.

4.4 NIQE mit speziell trainiertem Modell

Hier werden die Ergebnisse der NIQE-Metrik mit der Modellberechnung für eine aktuelle Videosequenz bzw. bestimmte Schicht von CT-Daten wie in Kapitel 2.3.3 beschrieben, diskutiert.

Für diese Auswertung sollen erneut die Videosequenz Crew bzw. die 10te Schicht der CT-Volumendaten Neck herangezogen werden. Hierbei wird also die Metrik gezielt auf die qua-

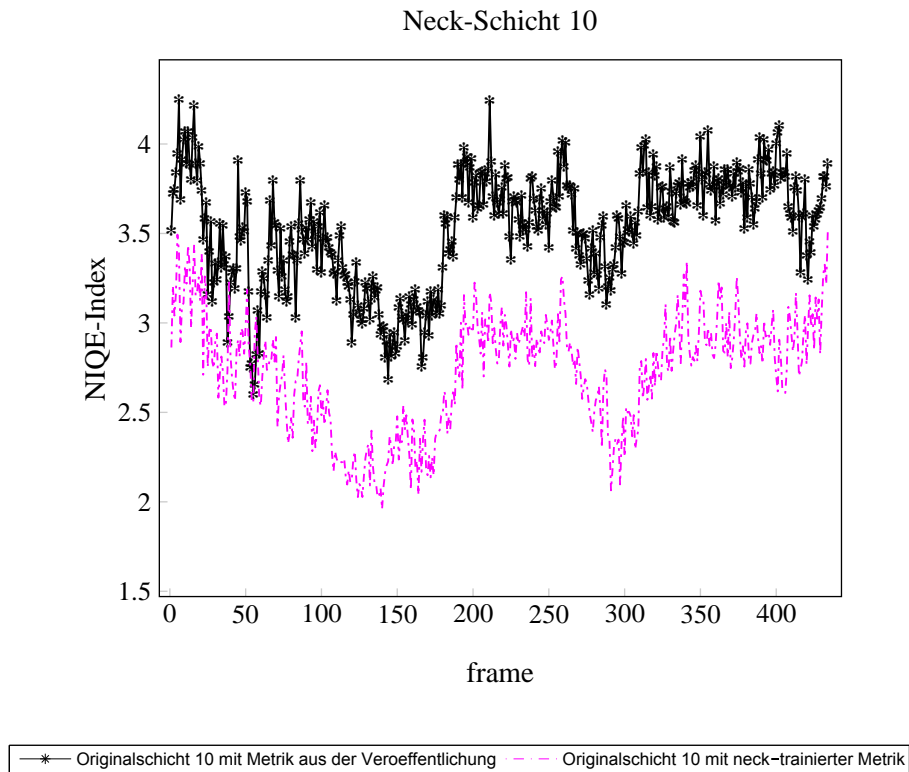


Abbildung 4.34: Ergebnisvergleich Neck-Schicht 10

litativ zu bewertenden Frames trainiert und deren besonderen Merkmale bereits bei der Modellgenerierung berücksichtigt.

4.4.1 Videosequenz Crew

4.4.1.1 Crew-Trainierter NIQE vs. PSNR - Vergleich Tiefpassbänder

Dem Plot für die crew-trainierte Metrik ist zunächst wenig zu entnehmen. Daher soll nun näher auf die einzelnen TPs eingegangen werden.

Betrachtet man die Ergebnisse für L-zero ist eine Tendenz der NIQE-Indexe zum Schlechteren wahrnehmbar. Weiter existieren sehr große Schwankungen, vgl. rote Kurve für L-zero in Abb. 4.35 im Bereich der TP-Stellen 101–189, in welchem viele schlechte NIQE-Indexe erzielt werden. Jedoch fehlt die Übereinstimmung, die Tendenz der NIQE-Indexe ausgenommen, der Extremwerte im Vergleich zum PSNR. Diese können, wie bereits in Kapitel 4.2.3.2. näher erläutert, allesamt mit den Blitzlichtern in Verbindung gebracht werden. Erwartungs-

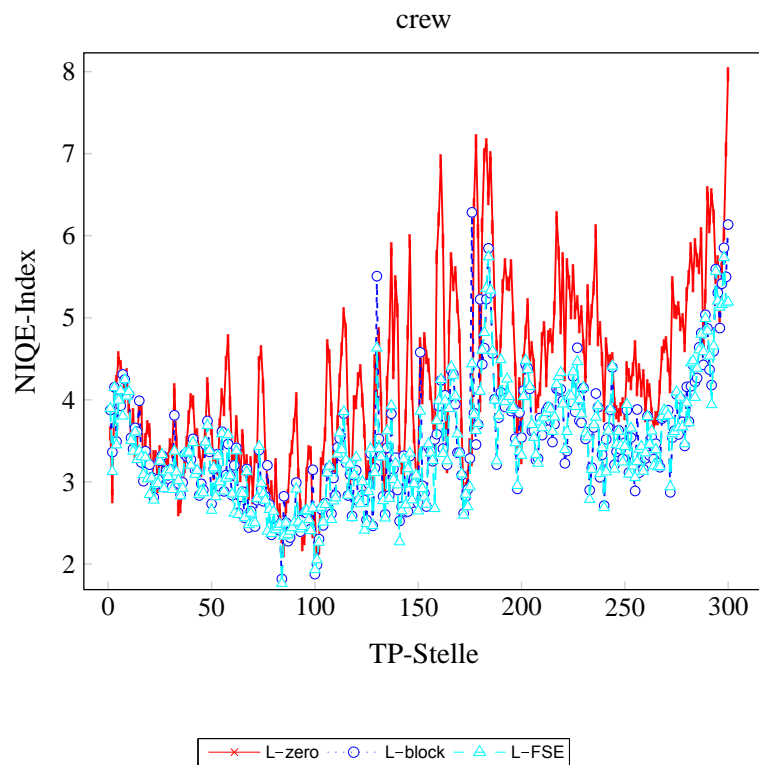


Abbildung 4.35: Ergebnisse der crew-trainierten NIQE-Metrik von der Crew-Sequenz

gemäß hätte sich der Kurvenverlauf von L-zero beim NIQE weit oberhalb der anderen TPs, wie es beim PSNR zu sehen ist, befinden müssen, es ist aber nur ein kleiner Abstand erkennbar.

Der Verlauf der blauen Kurve für L-block weist ebenso die Tendenz zu höheren NIQE-Indizes auf und ist extremen Schwankungen unterlegen. Dessen Extremwerte stimmen zwar nicht mit dem zuvor untersuchten L-zero überein, allerdings harmonisieren sie an wenigen TP-Stellen mit denen vom PSNR. Es handelt sich hierbei um die TP-Stellen 130, 151, 176 und 180, an denen für L-block bei beiden Metriken schlechte Ergebnisse resultieren, deren TP-Bilder starkes Blitzlicht enthalten. Dem gegenüber stehen jedoch wiederum die vom NIQE am besten bewerteten TPs. So werden z. B. an den TP-Stellen 84 und 100 sehr gute NIQE-Indexe erzielt, dies deckt sich aber nicht mit dem PSNR-Werten.

Die Ergebnisse für L-FSE entsprechen denen des L-blocks, beide TPs verhalten sich auch hier ähnlich und liegen auf nahezu gleicher Höhe und werden daher nicht näher untersucht.

Ursprüngliche Metrik vs. crew-trainierte Metrik	L-zero	L-block	L-FSE
Mittelwert NIQE-Index (Metrik aus der Veröffentlichung)	3.7246	3.5349	3.5646
Mittelwert NIQE-Index (crew-trainierte Metrik)	4.244	3.452	3.42

Tabelle 4.8: Mittelwertvergleich Crew-Sequenz

Dieses Ergebnis bricht viele Erwartungen an die neue Metrik und ist ein sehr wichtiges Indiz für die fehlende Eignung des NIQEs.

4.4.1.2 Crew-Trainierter NIQE vs. NIQE aus der Veröffentlichung

4.4.1.2.1 Vergleich Tiefpassbänder Bei dieser Untersuchung fällt erneut auf, dass an manchen TP-Stellen noch schlechtere NIQE-Indexe existieren, als mit der Metrik nach der Veröffentlichung vorhanden waren. Der NIQE-Indexbereich bei L-zero lag bei 2.9–4.9, nun befinden sich die NIQE-Indexe zwischen ca. 2 und 8. Für L-block bzw. L-FSE betrug der NIQE-Indexbereich 2.8–4.8, während die NIQE-Indexe mithilfe der crew-trainierten Metrik zwischen 1.8 und 6.3 liegen. Ein Grund hierfür könnte sein, dass die crew-trainierte Metrik nun ebenfalls auf die Blitzlichter sensibler reagiert und TP-Bilder ohne Blitzlichter nun besser bewertet, wodurch die extremen Schwankungen entstehen. Beim Vergleich der Mittelwerte, aufgeführt in Tab. 4.8, der crew-trainierten Metrik mit denen, der aus der Veröffentlichung stammenden Metrik, zeigt, dass sich dieser für L-zero um ca. 0.5 erhöht. Die anderen beiden TPs erfahren keine nennenswerte Verbesserung. Bei Betrachtung der Plots beider Metriken ist kein kurvenähnlicher Verlauf aller TPs auszumachen. Jedoch ändern sich die Kurven nun in Bezug auf die Tendenz, was mit dem Video in Einklang steht, wessen Qualität aus visueller Sicht zunehmend schlechter wird. Ansonsten liefert die crew-trainierte Metrik wider Erwarten keine verwertbaren Ergebnisse, da die PSNR-Werte im Vergleich hierzu hoch ausfallen, einhergehend mit der guten Qualität für diese Videosequenz. Dies ist jedoch selbst bei der crew-trainierten Metrik nicht eindeutig ersichtlich.

4.4.1.2.2 Vergleich Originalvideo Die bislang gewonnenen Erkenntnisse sprechen überwiegend gegen den NIQE und lassen dessen Eignung zur Qualitätsbewertung sehr bezwei-

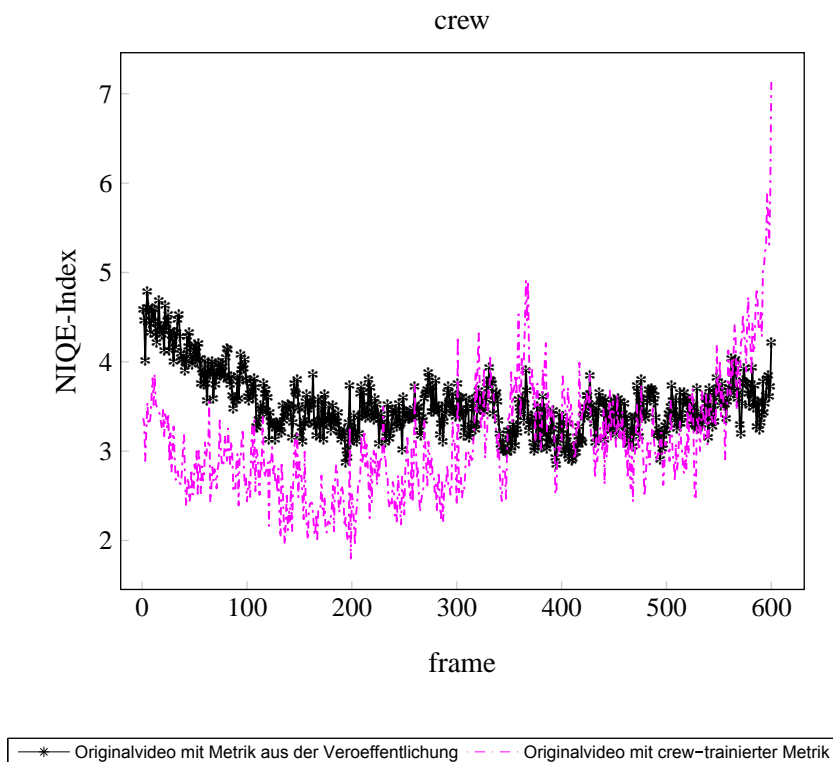


Abbildung 4.36: Ergebnisvergleich Crew-Sequenz

feln. Um dies jedoch zu festigen ist eine Prüfung des Originalvideos unabdingbar. Abermals sollen die Ergebnisse gegenübergestellt werden, die mit der aus der Veröffentlichung stammenden Metrik erzielt wurden, mit denen, die aus der crew-trainierten Metrik resultieren. Ein weiteres Indiz für die Untauglichkeit des NIQEs liefern die Ergebnisse der Frames von 331–427 und 548–600, siehe Abb. 4.36. Es stellt sich wiederum die Frage, wieso die Qualität für Frames plötzlich schlechter ausfällt, obwohl die Metrik speziell auf diese Frames trainiert wurde. Die NIQE-Indexe hätten für die crew-trainierte Metrik erwartungsgemäß nahe oder gleich 0 sein müssen, da es sich hierbei um das Original handelt, dessen PSNR-Werte beispielsweise auch nachvollziehbar im unendlichen Bereich befinden.

4.4.2 Schicht 10 der CT-Volumendaten Neck

Mit dieser Methode wird nochmals ein weitergehender Ansatz als in Kapitel 4.3.3.3 versucht, nachdem diese Ergebnisse zu keinen präzisen Aussagen geführt hatten. Hiermit soll

Neck-Schicht 10

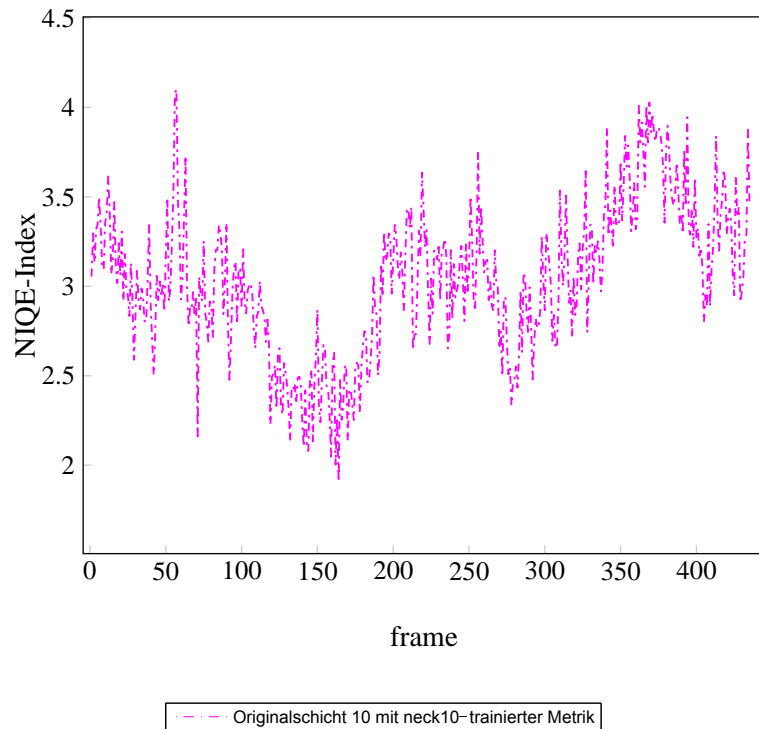


Abbildung 4.37: Ergebnis der neck10-trainierten NIQE-Metrik von der Neck-Schicht 10

die Eignungsfähigkeit der neuen Metrik explizit auf die Probe gestellt werden.

Zunächst soll auf die Erzeugung dieses Modells eingegangen werden. Prinzipiell handelt es sich um dieselbe Vorgehensweise wie unter Kapitel 4.3, hierbei wurde die Metrik jedoch auf eine spezielle Schicht der Neck-Sequenz trainiert. Es wurden alle Frames aus der 10ten Schicht dieser Sequenz herangezogen, um die Metrik darauf sowohl zu trainieren als auch zu testen. Demnach ist mit diesem Modell die Metrik auf ein und dieselben Frames trainiert sowie getestet worden, vgl. Kapitel 4.4.1 zur Videosequenz Crew, nur dass es sich hierbei um CT-Volumendaten handelt.

Diese Schicht wurde bereits mit der auf jedes dritte Frame trainierten Metrik der gesamten Neck-Sequenz ausgewertet, wobei zu erkennen war, dass eine deutliche Verbesserung der NIQE-Indexe folgte. Für die NIQE-Indexe nach der Metrik aus der Veröffentlichung ändert sich nichts, da die auszuwertenden Frames übereinstimmen, daher bleiben diese nun außen vor. Vergleicht man aber nun die NIQE-Indexe mit der speziell auf alle Frames der

10ten Schicht trainierten Metrik mit den Ergebnissen aus Kapitel 4.3.3.3 sind deutlich größere Schwankungen zu erkennen. Die NIQE-Indexe variieren hierbei von Frame zu Frame viel stärker, der Kurvenverlauf bleibt jedoch derselbe. Bei Untersuchung der herausragenden Stelle bei Frame 57, siehe Abb. 4.37, in Verbindung mit dem hierzu korrespondierenden Bild aus der Sequenz sind wiederum Zahnfüllungen zu erkennen, die jedoch nun den NIQE-Index negativ beeinflussen, der nun bei 4.102 liegt. Allerdings sind die Zahnfüllungen von Frame 54 bis 58 zu sehen, was plötzlich mit den Ergebnissen für diese Frames harmonisiert. Im Vergleich zur Abb. 4.34 kann nun die speziell trainierte Metrik darauf sensibler reagieren und liefert nachvollziehbarere Ergebnisse.

Der NIQE-Indexbereich befindet sich jetzt bei 1.909–4.102 und ist damit schlechter als der in Kapitel 4.3.3.3 ermittelte. Ebenso bedeutsam ist die Verschlechterung des Mittelwertes auf 3.035. Einerseits hätten der NIQE-Indexbereich sowie der Mittelwert erheblich niedriger liegen müssen. Es bleibt dennoch zu berücksichtigen, dass sich durch spezielles Training der Metrik die visuelle Bildqualität der auszuwertenden Frames nicht verbessert. Ein Beispiel für diesen Ansatz liefert erneut Frame 57. An dieser Stelle ist der NIQE-Index höher als vorher, allerdings kann dies nun nachvollzogen werden. Nach Kapitel 4.3.3.3 war dieses Frame nicht sonderlich herausragend. Nun aber ist deutlich festzuhalten, dass die Zahnfüllung das Bild etwas verzerrt und vermutlich deshalb der NIQE-Index schlechter ausfällt. Selbst wenn die Metrik auf alle Frames hierauf trainiert wurde, kann wohl letzten Endes nicht erwartet werden, dass die NIQE-Indexe gegen 0 gehen, auch nicht für ein Originalbild.

Kapitel 5

Zusammenfassung der Ergebnisse zur neuen Metrik

Bei der Basketball-Sequenz handelt es sich um ein Video, das reich an Bewegungen ist, wie Positionsveränderungen der Bildobjekte, die Kameradrehung sowie das Zoom-In bzw. Zoom-Out. Dies führt wiederum zu unscharfen L-zeros und zu L-blocks bzw. L-FSEs mit Artefakten und zudem verschwommenen Stellen in den jeweiligen TP-Bildern. Die visuelle Qualität dieser geht jedoch nicht mit den zugehörigen NIQE-Indexen einher. Dafür dass das L-zero stets unscharf abgebildet ist und L-block bzw. L-FSE neben einigen verschwommenen Stellen darüber hinaus immerzu Artefakte aufweisen, sind die NIQE-Indexe zu starken Schwankungen unterlegen. Ebenso wechselt die Rangfolge der TPs untereinander zu häufig, obwohl durchgehend dieselben Störungen auszumachen sind. Was die R-origs anbelangt, so entsprechen die Ergebnisse nicht der Erwartung. Zumal es sich hierbei um die Originalbilder handelt, fallen die NIQE-Indexe nicht niedrig genug aus. Insofern ist kein System erkennbar, wie sich Unschärfe und Artefakte auf die NIQE-Metrik auswirken.

Auch die Discovery-City-Sequenz enthält viele Schnellaufnahmen und teilt sich in kleine Teilsequenzen auf. Am Plot ist ersichtlich, wann ein Kurzvideo aufhört und das nächste beginnt, da sich die NIQE-Indexe für alle TPs sowie des R-origs tendenziell gleich ändern. Demnach muss es einen Zusammenhang der TP-Bilder mit den NIQE-Indexen geben. Und

dennoch können die meisten Ergebnisse nicht begründet werden. So z. B. an Transitionen, an welchen das L-zero durchgehend 2 Frames ineinander deutlich wiedergibt, aber der NIQE-Index stets schlecht bzw. am schlechtesten ausfällt, während die beiden kompensierten TPs mit variierenden NIQE-Indizes beurteilt werden. Ebenso fragwürdig ist hierbei, weshalb auch das R-orig an diesen Übergängen immerzu schlecht bzw. am schlechtesten bewertet wird, zumal es sich um das zugehörige R-orig handelt. Darüber hinaus ist erneut nicht ersichtlich, inwieweit Unschärfe bzw. Artefakte für den NIQE eine Rolle spielen.

Das Besondere der Crew-Sequenz sind die Blitzlichter, welche einen wachsenden Einfluss auf die visuelle Qualität der Bilder, aufgrund zunehmender Lichtreflexion, haben, je näher sich die Astronauten an der Kamera befinden. Sobald ein Frame ein Blitzlicht aufweist, enthält das zugehörige TP umso mehr Artefakte, hinsichtlich des stärkeren Kontrasts. Darüber hinaus nimmt die Schärfe des L-zeros kontinuierlich ab, was jedoch nicht mit den NIQE-Indizes korrespondiert. Wie bereits erläutert wirkt sich der Kontrast auf die Varianz σ aus, dies wiederum überträgt sich auf das unter Kapitel 2.2.4 beschriebene MVG-Modell. Demnach hätten sich NIQE-Indizes erwartungsgemäß auffallend schlecht für solche TPs ergeben müssen. Ebenso zeigt sich die Metrik unbeeindruckt, was die Unschärfe bzw. die Artefakte angeht, da die NIQE-Indizes für eine grundlegende Aussage zu sehr variieren.

Für die Flower-Sequenz ergeben sich keine Besonderheiten, die hätten näher untersucht werden können. Im Laufe des Videos erfährt lediglich die Kamera eine gemächliche Drehung und anhand des NIQE-Indexbereichs ist erkennbar, dass sich auch diese in einem bestimmten Rahmen befinden. Positiv auffallend ist, dass das stets unscharfe L-zero ebenso von der Metrik nahezu durchgehend für qualitativ schlechter als die kompensierten TPs befunden wird. Allerdings ist das beständige Auf und Ab der Kurven aller TPs sowie des R-origs nicht verwertbar und verhalten sich eher unkonventionell zu den zugehörigen Bildern.

Ausschlaggebend bei der Foreman-Sequenz ist die Kameradrehung. Während dieser sind alle TP-Bilder sowie das R-orig völlig unscharf abgebildet. Diese Szene korrespondiert mit den NIQE-Indizes, welche sich abrupt für alle TPs sowie R-origs massiv erhöhen, und kann als positiver Aspekt der neuen Metrik gewertet werden. Es kann davon ausgegangen werden, dass der NIQE gravierende Unschärfe entsprechend berücksichtigt. Dennoch kann keine In-

terpretation der NIQE-Indexe und deren genaue Bedeutung angestellt werden.

Die Einflussnahme des Rauschens auf den NIQE ist zu Beginn der Discovery-Orient-Sequenz gegensätzlich zu dem, wie es aus der Foreman-Sequenz hervorgeht. Es wurden Beispiele gefunden, welche die Vermutung bezüglich des Rauschens anzweifeln lassen, dass der Rauschanteil beim R-orig aufgrund einer Mittelung größer ist als beim L-zero und daher einen negativen Einfluss auf die Metrik habe und deshalb schlechter bewertet werde. Dennoch war ganz klar stärkeres Schwanken der NIQE-Indexe bei Teilsequenzen mit Rauschen zu vernehmen. Ähnlich der Discovery-City-Sequenz ist an den Plots ebenso deutlich zu erkennen, wann ein Kurzvideo endet und das nächste beginnt, da sich auch hier die NIQE-Indexe für alle TPs sowie R-origs tendenziell gleich verändern. Jedoch fehlt nach wie vor die Bedeutung der NIQE-Indexe zur Begründung der Ergebnisse in Bezug auf die TP-Bilder und trotzdem scheint es einen Aspekt zu geben, der die Metrik überwiegend beeinflusst. Dass die NIQE-Indexe für verschiedene Teilaufnahmen variieren und sich mit ihren NIQE-Indexen zu anderen Kurzaufnahmen differenzieren ist definitiv kein Zufall.

Entgegen aller Erwartungen fielen die Ergebnisse der Metrik aus der Veröffentlichung bei Anwendung auf CT-Volumendaten aus. Die Schwankungen, Rangfolge der TPs und der NIQE-Indexbereich differenzieren sich nicht im Geringsten zu den Auswertungen bei den Videosequenzen. Die Untersuchung des NIQEs hinsichtlich dessen Eignungsfähigkeit bei Anwendung von unnatürlichen Bildern liefert diesbezüglich keine verwertbaren Ergebnisse. Ist der NIQE auf CT-Daten trainiert, ergibt sich nur bedingt eine Qualitätsverbesserung. Demnach sind die NIQE-Indexe einzelner Schichten offensichtlich von deren Größe, also der Anzahl der jeweiligen Frames, abhängig. Je größer die Schichten sind, umso stärker ist die Gewichtung bei der trainierten Metrik. Für diese Schichten verkleinern sich die NIQE-Indexe dennoch nur örtlich, wie das Beispiel der 4ten Schicht der Thorax-Sequenz zeigt. Des Weiteren existieren immer wieder Extremwerte, wonach die trainierte Metrik höhere NIQE-Indexe erzielt als die Metrik aus der Veröffentlichung und das ohne ersichtlichen Grund. Allerdings scheint die trainierte Metrik Zahnfüllungen besser interpretieren zu können. Beim Training und Testen des NIQEs auf ein und dieselben Frames von Videosequenzen sowie CT-Daten verhalten sich die NIQE-Indexe etwas besser nachvollziehbar. Trotzdem

lassen sich die NIQE-Indexe, speziell der crew-trainierten Metrik, nach wie vor nicht eindeutig auf die PSNR-Werte projizieren, sodass letztendlich auch eine Interpretation der NIQE-Indexe, also die Relation dieser mit den Bildern, nicht hergeleitet werden kann.

Generell konnten für alle verwendeten Sequenzen keine übereinstimmenden Ergebnisse bezüglich Artefakte, Rauschen bzw. Unschärfe ausfindig gemacht werden. Dies bedeutet wiederum, dass die Metrik keine verlässlichen Ergebnisse erzielt und die Qualitätsbewertung anhand des NIQEs nicht weiter zu empfehlen ist. Ein Indiz hierfür stellt bei allen Sequenzen das R-orig dar. Es befindet sich beim PSNR durchgehend im unendlichen Bereich, während es vom NIQE den TPs qualitativ gleichgestellt wird. Demnach muss die Vermutung angestellt werden, dass es einen NIQE-Index von 0 möglicherweise gar nicht gibt. Jedoch kann für diese Arbeit keine präzise Aussage zum bestmöglichen NIQE-Index getroffen werden.

Die Ergebnisse bezüglich einer neuen OU-DU IQA Metrik, welche die Qualität von Bildern blind bestimmen kann, führen nicht zum gewünschten Ergebnis. Die Tauglichkeit, vor allem bezüglich der TPs und deren Störungen, muss eindeutig in Frage gestellt werden. An den NIQE-Indexen war während der gesamten Auswertungen kein roter Faden bezüglich der Artefakte oder Unschärfe erkennbar. Infolgedessen kann das PSNR keineswegs durch den NIQE hinsichtlich der qualitativen Beurteilung von TPs ersetzt werden, nachdem sich dieser von Artefakten ebenso unbeeindruckt zeigt. Es wurden verschiedene Arten versucht, die NIQE-Metrik für das vorliegende Problem anzupassen. Es hat sich herausgestellt, dass die NIQE-Metrik für die Bewertung des Tiefpassbandes nicht sinnvoll verwendet werden kann. Die Untersuchung der Skalierbarkeit bezüglich der Rate wurde daher nicht näher geprüft, zumal sich diesbezüglich bereits Untersuchungen in der Veröffentlichung [MSB13] befinden. Ein möglicher Ansatz zur weiteren Prüfung könnte die Anwendung des PSNRs speziell auf Bereiche mit Update von den Update Frames sein. Die PSNR-Werte könnten erneut mit den NIQE-Indexen verglichen werden um evtl. Kenntnisse in Bezug auf die Berücksichtigung von Artefakten beim NIQE zu deuten. Abermals könnten Videosequenzen sowie CT-Daten herangezogen werden. Evtl. wäre hiermit sogar eine Interpretation der NIQE-Indexe möglich.

Abbildungsverzeichnis

2.1	Patch Selection	5
2.2	NIQE-Testbilder	8
3.1	L-block mit Artefakte	11
3.2	BSB L-zero	12
3.3	BSB L-block	12
3.4	Pixel	14
3.5	BSB L-FSE	14
4.1	Bsp. Basketball-Sequenz	17
4.2	Bsp. Discovery-City-Sequenz	17
4.3	Bsp. Crew-Sequenz	18
4.4	Bsp. Flower-Sequenz	19
4.5	Bsp. Foreman-Sequenz	20
4.6	Bsp. Discovery-Orient-Sequenz	20
4.7	NIQE-Indexe Basketball	22
4.8	PSNR-Werte Basketball	24
4.9	NIQE-Indexe Discovery-City	27
4.10	Discovery-City-TP-Stelle 22	29
4.11	PSNR-Werte Discovery-City	31
4.12	NIQE-Indexe Crew	34
4.13	Crew-TP-Stelle 20	35
4.14	Crew-TP-Stelle 300	38

4.15 Crew-TP-Stelle 209	39
4.16 PSNR-Werte Crew	40
4.17 NIQE-Indexe Flower	45
4.18 PSNR-Werte Flower	47
4.19 NIQE-Indexe Foreman	49
4.20 Foreman-TP-Stelle 33	50
4.21 PSNR-Werte Foreman	52
4.22 NIQE-Indexe Discovery-Orient	56
4.23 PSNR-Werte Discovery-Orient	59
4.24 NIQE-Indexe Head-Sequenz	62
4.25 PSNR-Werte Head-Sequenz	63
4.26 head-trainiert	66
4.27 Schicht0-head-trainiert	67
4.28 Schicht4-head-trainiert	68
4.29 thorax-trainiert	69
4.30 Schicht1-thorax-trainiert	70
4.31 Schicht4-thorax-trainiert	71
4.32 neck-trainiert	72
4.33 Schicht1-neck-trainiert	73
4.34 Schicht10-neck-trainiert	75
4.35 TP-NIQE-Indexe crew-trainiert	76
4.36 Original-NIQE-Indexe crew-trainiert	78
4.37 NIQE-Indexe neck10-trainiert	79

Tabellenverzeichnis

4.1	Mittelwerte Basketball-Sequenz	25
4.2	Mittelwerte Discovery-City-Sequenz	32
4.3	Mittelwerte Crew-Sequenz	42
4.4	Mittelwerte Flower-Sequenz	47
4.5	Mittelwerte Foreman-Sequenz	54
4.6	Mittelwerte Discovery-Orient-Sequenz	60
4.7	Mittelwerte Head-Sequenz	64
4.8	Mittelwertvergleich Crew	77

Literaturverzeichnis

- [CDSY97] A.R. Calderbank, I. Daubechies, W. Sweldens, and Boon-Lock Yeo. Lossless image compression using integer to integer wavelet transforms. In *Image Processing, 1997. Proceedings., International Conference on*, volume 1, pages 596–599 vol.1, 1997.
- [GH05] B. Girod and Sangeun Han. Optimum update for motion-compensated lifting. *Signal Processing Letters, IEEE*, 12(2):150–153, 2005.
- [Kau12] Prof. Dr.-Ing. A. Kaup. Image and vido compression, Summer Term 2012.
- [MSB13] A. Mittal, R. Soundararajan, and A.C. Bovik. Making a completely blind image quality analyzer. *Signal Processing Letters, IEEE*, 20(3):209–212, 2013.
- [SK10] J. Seiler and A. Kaup. Complex-Valued Frequency Selective Extrapolation for Fast Image and Video Signal Extrapolation. *IEEE*, 17(11):949–952, November 2010.
- [SSK13] W. Schnurrer, J. Seiler, and A. Kaup. Improving block-based compensated wavelet lifting by reconstructing unconnected pixels. In *International Symposium on Signals, Circuits and Systems (ISSCS)*, Iasi, Romania, July 2013.