Impact of Pooling Methods on Image Quality Metrics

David Norman Díaz Estrada and Marius Pedersen
Department of Computer Science, *Norwegian University of Science and Technology*, Gjøvik, Norway
davidnd@stud.ntnu.no, marius.pedersen@ntnu.no

***Abstract*—Image quality assessment using objective metrics is becoming more widespread, and an impressive number of image quality metrics have been proposed in the literature. An aspect that has received little attention compared to the design of these metrics is pooling. In pooling the quality values, usually from every pixel is reduced to fewer numbers, usually a single value, that represents overall quality. In this paper we investigate the impact of different pooling techniques on the performance of image quality metrics. We have tested different pooling methods with the SSIM and S-CIELAB image quality metrics in the CID:IQ database, and found that the pooling technique has a significant impact on their performance.**

***Index Terms*—image quality, metrics, pooling**

I. INTRODUCTION

In the field of Image Quality (IQ) assessment, there are two main branches: subjective and objective. In the branch of subjective IQ assessment the Mean Opinion Score (MOS) is based on scores given by human observers and is considered as the ground truth, on the other hand, in objective IQ assessment the goal is to use IQ metrics which are computational methods that predict the IQ without the need of human observers. For the objective assessment there are different types of IQ metrics: Full-Reference, No-Reference, and Reduced-Reference [8] depending on the availability of the reference image. Many IQ metrics have been proposed in the literature [8], including extensive evaluations of them [7, 8].

Most IQ metrics follow a similar design [8], normally they would produce a quality value for each pixel in the image, producing a quality map. The last step of most IQ metrics is to carry out pooling of the quality map. Pooling is applied to summarize all values in the quality map into a single value [2, 15]. In the literature, the most common pooling method used is the spatial average, however, more advanced pooling methods can lead to higher correlation with MOS [1, 2, 6, 11, 12, 14, 15]. Many contributions have been made in designing IQ metrics, however, pooling has been studied to a lesser extent. The goal of the paper is to investigate the influence of pooling methods on the performance of IQ metrics.

The contributions of this paper can be summarized as:

- Analysis of pooling methods on the performance of two state-of-the-art IQ metrics, S-CIELAB [16] and SSIM [13], on the CID:IQ Database.
- Proposal for a modification of the five number summary pooling technique.

In this paper, Section II contains a brief overview of some previous studies that have evaluated pooling techniques, and presents the existing quality based pooling methods. Section III describes the experimental setup, the dataset and quality metrics used. Section IV presents the results of this project. Later on, Section V concerns the discussion of results. Lastly, section VI gives an overview of the results and concludes the paper along with some ideas about possible future work.

II. BACKGROUND

There are two main branches in pooling methods: content-based and quality-based. In content-based pooling the objective is to identify the most relevant regions in the image, for example by using information of a saliency map to weight the importance of each element in the quality map provided by the objective metric [2, 15]. On the other hand, quality-based pooling relies only on the quality map itself, without the need of additional content information. In this paper the main focus is on quality-based pooling methods.

A. Related studies

In the literature, the most recent studies that have carried out evaluations of different pooling methods are the following:

Wang and Shang (2006) [14] evaluated three pooling strategies on the LIVE database [11], and concluded that all improved the performance of the SSIM metric compared to taking the spatial average.

Pedersen and Gong (2012) [2] studied the impact of pooling methods on color printing quality attributes by evaluating six IQ metrics and six pooling methods. They showed that pooling is metric dependent, and that more advanced pooling metrics generally achieve higher correlation than mean pooling [2].

Zewdie et al. (2014) [15] proposed the Five Number Summary (FNS) as a pooling strategy for SSIM and compared it against four other pooling methods, which were evaluated on TID2008 [10], TID2013 [9] and LIVE release 2 [3] databases. They showed that the FNS generally gives higher correlation than other pooling methods.

Temel and AlRegib (2015) [12] evaluated three IQ metrics and 8 pooling methods on LIVE [11], multiply distorted LIVE

David Norman Díaz Estrada is supported by the Erasmus+ Joint Master's Degree Computational Colour and Spectral Imaging (COSI). Marius Pedersen is supported by the project Quality and Content (project number 324663) funded by the Research Council of Norway.

978-1-7281-8750-1/20/$31.00 ©2022 IEEE

[4] and TID2013 [9] databases. They also proposed a Weighted Percentile Pooling (WPP). This study concluded that mean pooling was less consistent than other pooling strategies in terms of variations along different databases, distortion types, and quality attributes [12].

B. Quality based pooling methods

The most common way of pooling is done by taking the spatial average, moreover, other basic statistics can also be used to summarize the quality map, such as the minimum (min), maximum (max), median, Q1 (first quartile), Q3 (third quartile), 95 percentile (P95) and the standard deviation (STD). However, more advanced spatial pooling methods follow a general form [2, 15]:

$$M = \frac{\sum_{i=1}^{n} w_i m_i}{\sum_{i=1}^{n} w_i} \tag{1}$$

where m_i represents the quality value at location i, and w_i is the correspondent weight assigned to that location, and n is the total number of values in the quality map.

1) Minkowski pooling: In Minkowski pooling [1] the strategy consists of raising each value in the quality map to a given power p and then computing the average of those values. The most common values used in the literature are $p = [1/8, 1/4, 1/2, 2, 4, 8]$. In this pooling method, the power p affects the quality map by making higher distortion values (when the IQ metric outputs a distortion map) or lower quality values (when the metric outputs a quality map, i.e. SSIM [13]) to have a larger effect:

$$M = \frac{1}{N} \sum_{i=1}^{n} {m_i}^p \tag{2}$$

When $p = 1$ it is the mean absolute error, while a $p = 2$ becomes the mean squared error.

2) Monotonic Function pooling: Monotonic function pooling [14] takes the form of a weighted average, where the weight w_i is computed as $|m_i|^p$. Moreover, when the IQ metric outputs a quality map the values of p are negative powers (monotonically decreasing) where the most common values used in the literature are $p = [-1/8, -1/4, -1/2, -1-2, -4, -8]$; conversely, if the metric outputs a distortion map, then a monotonically increasing function is used (with positive values of p):

$$M = \frac{\sum_{i=1}^{n} |m_i|^p m_i}{\sum_{i=1}^{n} |m_i|^p} \tag{3}$$

3) Percentile pooling: Percentile pooling [6] is motivated by the idea that the lowest quality values affect the perceived IQ in a non-proportional manner, and thus, they should receive a heavier weighting. Percentile pooling is computed as in Equation (4), where for a given distortion value m_i, the corresponding weight w_i is determined by a step function $f(m_i)$ as in Equation (5), where T is an adaptive threshold, and r is a scaling factor; typically $T = 6^{th}$ percentile and $r = 1.1$ [6].

$$M = \frac{\sum_{i=1}^{n} f(m_i) m_i}{\sum_{i=1}^{n} f(m_i)} \tag{4}$$

$$w_i = f(m_i) = \begin{cases} r: & m_i > T \\ 1: & m_i < T \end{cases} \tag{5}$$

4) Five Number Summary: The Five Number Summary (FNS) as a pooling method was proposed in [15]. This method uses five representative values from the quality map: min, Q1 (the first quartile, equivalent to the 25^{th} percentile), median, Q3 (the third quartile, equivalent to the 75^{th} percentile) and max, and finally the average of these values is computed. The advantage of this method over other pooling strategies is that it only relies on the statistics of the quality map, and no weights are needed:

$$FNS = \frac{min + Q1 + median + Q3 + max}{5} \tag{6}$$

III. EXPERIMENTAL SETUP

A. Pooling techniques

We include the following techniques; mean, max, min, Q1, median, Q3, 95^{th} percentile (P95), Minkowski pooling, monotonic function pooling, percentile pooling, and FNS. Both for Minkowski and Monotonic Function pooling we use the following p values: {1/8,1/4,1/2,2,4,8} and p:{ -1/8,-1/4,-1/2,-1,-2,-4,-8}, respectively, since these are the most common values used in the literature. For Percentile Pooling we set T = 6^{th} percentile and r = 1.1, similarly as in [6, 15].

For FNS we would like to test different combinations as the traditional five number summary might not be the optimal configuration. Therefore, we test the following six combinations. FNS$_1$ which is the original FNS:

$$FNS_1 = \frac{min + Q1 + median + Q3 + max}{5} \tag{7}$$

FNS$_2$ that includes the mean:

$$FNS_2 = \frac{min + Q1 + median + Q3 + max + mean}{6} \tag{8}$$

FNS$_3$ that includes the mean instead of the minimum:

$$FNS_3 = \frac{mean + Q1 + median + Q3 + max}{5} \tag{9}$$

FNS$_4$ that includes the mean and P95 (95^{th} percentile) instead of the minimum and maximum:

$$FNS_4 = \frac{mean + Q1 + median + Q3 + P95}{5} \tag{10}$$

FNS$_5$ that includes the mean instead of median and maximum:

$$FNS_5 = \frac{min + Q1 + mean + Q3}{4} \tag{11}$$

FNS$_6$ that can be used to weight (λ is a weight between 0 and 1) the importance of low quality and high quality values.

$$FNS_6 = \frac{\lambda(Q1 + median) + mean + (1 - \lambda)(Q3 + P95)}{5}. \tag{12}$$

B. Dataset

The CID:IQ dataset [5] was used to test the different pooling methods. The dataset contains 23 reference images, along with a total of 690 distorted images corresponding to 6 types of distortions applied at 5 different levels. The distortions include: JPEG compression, JPEG2000 compression, Poisson noise, Gaussian blur, DeltaE gamut mapping and SGCK gamut mapping. The MOS included in the dataset corresponds to 17 observers and 2 viewing distances (50 cm and 100 cm).

C. Image quality metrics

The Structural Similarity Index (SSIM) [13] was selected because of it's widespread use, and that it produces a quality map with quality values for each pixel. SSIM measures similarity between a reference and a test image, where the output values in the quality map are in range [-1,1]. The other IQ metric is the S-CIELAB [16], which is a spatial extension of the well-known color difference metrics ΔE^*_{ab}. S-CIELAB calculates the image difference between an original and a reproduction. For S-CIELAB Minkowski p-values of 4 and 8 have not been used.

D. Performance measures

For measuring the performance of each pooling method we use correlation measures with non-linear fitting. For each of the pooling strategies, the Spearman, Pearson, and Kendall correlation coefficients between the subjective score (MOS) and the objective metric are reported. Moreover, since the MOS and the objective scores given by the metrics usually do not follow a linear relationship, it was necessary to adjust the non-linearity of the data, therefore, similarly as in [6, 11, 15] the following logistic function was used to tackle this issue:

$$Quality(x) = \beta_1 logistic(\beta_2, (x - \beta_3)) + \beta_4 x + \beta_5 \quad (13)$$

where x corresponds to the objective score given by the metric, and the variables are initialized as; β_1 is the max of MOS, β_2 is the minimum of MOS, β_3 is the mean of the objective scores and $\beta_4 = \beta_5 = 0.1$. Furthermore, the logistic function is given as follows:

$$logistic(\tau, x) = \frac{1}{2} - \frac{1}{1 + exp(\tau x)}. \quad (14)$$

IV. RESULTS

Figures 1 - 4 show the Pearson correlation plot for SSIM and S-CIELAB with 95% confidence intervals for 50cm and 100cm for each of the pooling methods. Overall we can see that many techniques have a similar performance, but with some techniques performing worse compared to the best.

Table I and Table II shows the summary of results with the Spearman, Pearson and Kendall correlations for each pooling method for SSIM for 50cm and 100cm, respectively. Looking at the Pearson correlation for 50 cm we see that the highest value is obtained with Minkowski pooling with $p = 8$, while the highest Spearman correlation is with Minkowski pooling with $p = 8$ and FNS6 (with $\lambda = 0.8$,$\lambda = 0.9$,$\lambda = 1$), and for Kendall the highest values are with FNS4 and FNS6 (with $\lambda = 0.8$,$\lambda = 0.9$,$\lambda = 1$. For 100cm mean, percentile pooling, Monotonic ($p = 1/8, 1/4 and 1/2$) and minkowski ($p = 2$) has the highest Pearson correlation value, the highest Spearman values are found with Minkowski ($p = 2$), Monotonic ($p = 1/8, 1/4 and 1/2$) and Percentile pooling, and for Kendall correlation the highest values are for Minkowski ($p = 2$), Monotonic ($p = 1/8, 1/4 and 1/2$) and Percentile pooling. Barplots of Pearson values are shown in Figures 1 and 2.

TABLE I
SUMMARY OF RESULTS (SSIM AND 50 CM). SPEARMAN, PEARSON AND KENDALL CORRELATIONS AFTER NON-LINEAR REGRESSION.

Pooling Method		**CIDIQ Database (50 cm)**		
	Parameter	Pearson	Spearman	Kendall
Mean		0.767	0.770	0.579
Max		0.258	0.460	0.350
Min		0.422	0.501	0.374
Q1		0.776	0.777	0.590
Median		0.726	0.723	0.550
Q3		0.624	0.606	0.457
P95		0.554	0.539	0.409
Minkowski	p=1/8	0.757	0.760	0.569
	p=1/4	0.759	0.762	0.571
	p=1/2	0.763	0.765	0.574
	p=2	0.775	0.777	0.585
	p=4	0.684	0.785	0.594
	p=8	0.790	0.790	0.601
Percentile Pooling		0.770	0.772	0.581
Monotonic	p=1/8	0.770	0.772	0.581
	p=1/4	0.772	0.774	0.583
	p=1/2	0.775	0.778	0.587
	p=1	0.780	0.783	0.592
	p=2	0.785	0.787	0.597
	p=4	0.784	0.785	0.597
	p=8	0.767	0.766	0.582
FNS1		0.539	0.576	0.428
FNS2		0.622	0.593	0.440
FNS3		0.787	0.789	0.601
FNS4		0.787	0.789	0.602
FNS5		0.539	0.576	0.428
FNS6	$\lambda = 0$	0.774	0.778	0.590
FNS6	$\lambda = 0.1$	0.778	0.782	0.593
FNS6	$\lambda = 0.2$	0.781	0.784	0.595
FNS6	$\lambda = 0.3$	0.783	0.786	0.597
FNS6	$\lambda = 0.4$	0.784	0.787	0.599
FNS6	$\lambda = 0.5$	0.786	0.788	0.600
FNS6	$\lambda = 0.6$	0.786	0.789	0.600
FNS6	$\lambda = 0.7$	0.787	0.789	0.601
FNS6	$\lambda = 0.8$	0.787	0.790	0.602
FNS6	$\lambda = 0.9$	0.788	0.790	0.602
FNS6	$\lambda = 1$	0.788	0.790	0.602

Table III and Table IV shows the summary of results with the Spearman, Pearson and Kendall correlations for each pooling method for S-CIELAB for 50cm and 100cm, respectively. For both 50cm and 100, the highest Pearson, Spearman and Kendall correlation is found with P95. Barplots of Pearson correlation values are shown in Figures 3 and 4.

V. DISCUSSION

From the results in Tables I-IV we can see there is not a single pooling technique that performs best for the two

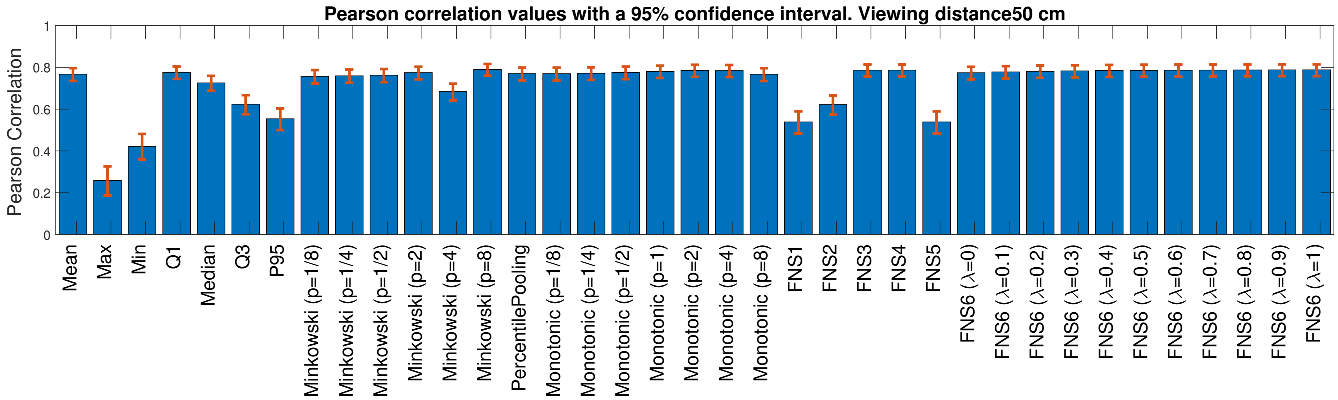

Fig. 1. Pearson Correlations with 95% confidence intervals for each pooling method for SSIM and 50 cm.

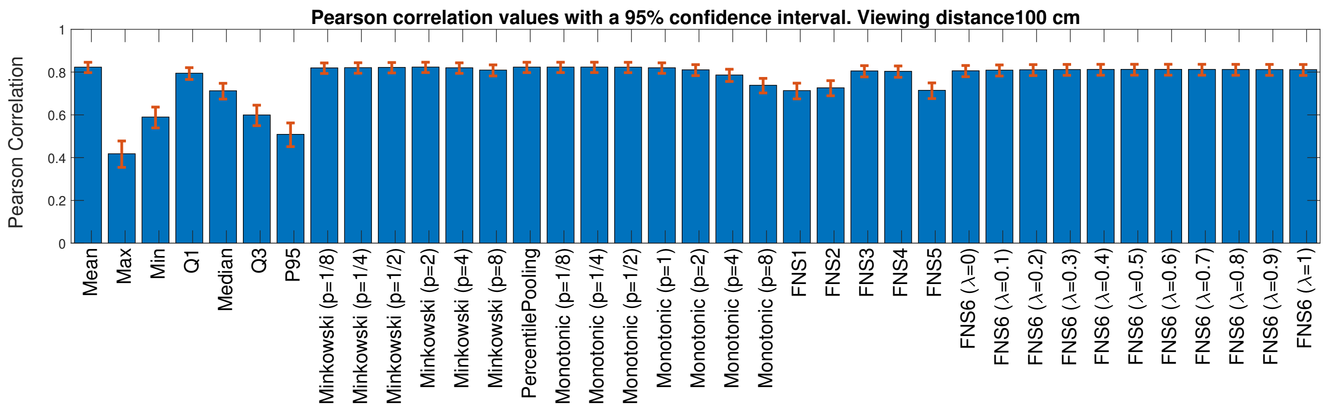

Fig. 2. Pearson Correlations with 95% confidence intervals for each pooling method for SSIM and 100 cm.

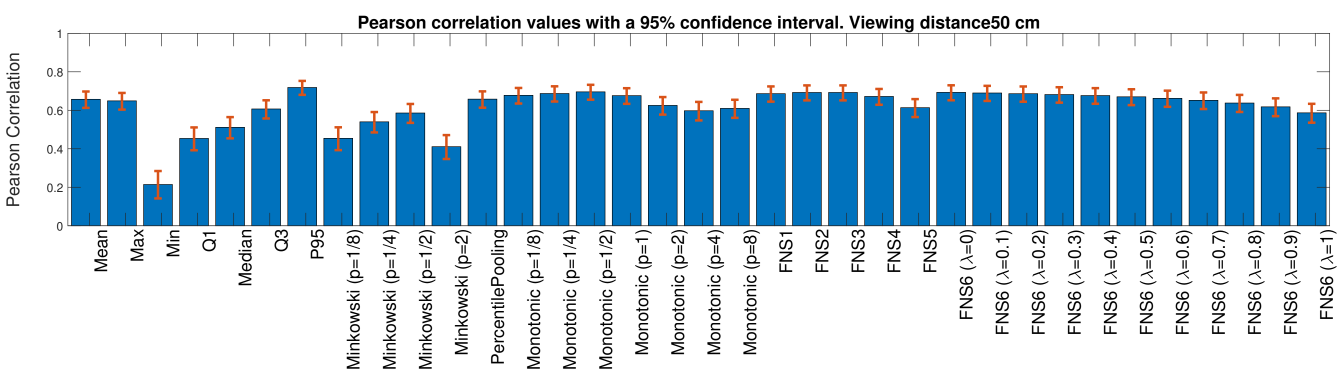

Fig. 3. Pearson Correlations with 95% confidence intervals for each pooling method for S-CIELAB and 50 cm.

image quality metrics and for both viewing distances. We can notice that mean pooling is not always the best performing, although it is commonly used. For S-CIELAB it is statistically significantly worse than P95.

We can notice that for the suggested FNS6, SSIM seems to be more stable when it comes to the λ parameter compared to S-CIELAB. We can also see that FNS3, FNS4 and FNS6 for SSIM performs similar to the best pooling techniques, which can also be found for S-CIELAB. This indicates that the suggested pooling techniques are stable across the metrics.

VI. CONCLUSIONS AND FUTURE WORK

The main contributions of this paper include the analysis of different pooling methods. They have been evaluated using SSIM and S-CIELAB on the CID:IQ Database. Moreover, another contribution is modifications for the Five Number Summary. The results show that the Five Number Summary (FNS) and its modifications achieve high correlations. The results indicate also that a single pooling technique is not able to achieve the highest performance for the two image quality

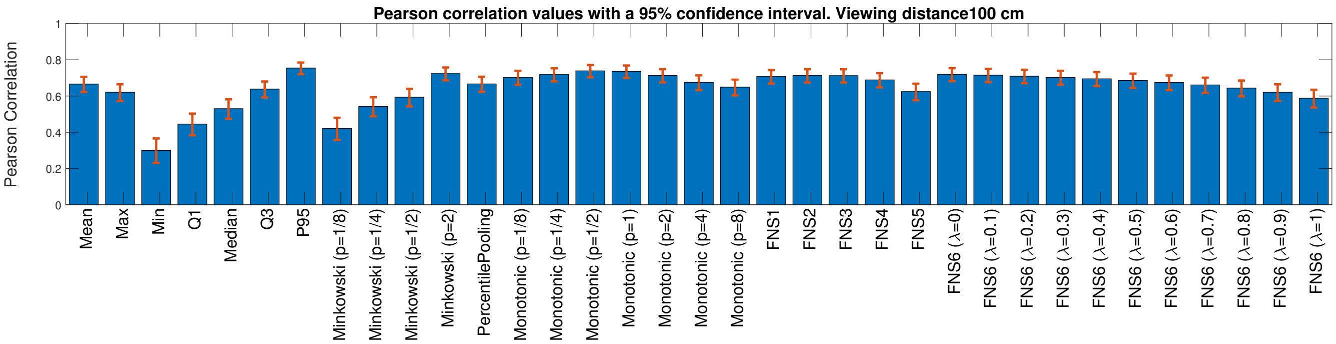

Fig. 4. Pearson Correlations with 95% confidence intervals for each pooling method for S-CIELAB and 100 cm.

TABLE II
SUMMARY OF RESULTS (SSIM AND 100 CM). SPEARMAN, PEARSON AND KENDALL CORRELATIONS AFTER NON-LINEAR REGRESSION.

Pooling Method		**CIDIQ Database (100 cm)**		
	Parameter	Pearson	Spearman	Kendall
Mean		0.823	0.809	0.612
Max		0.418	0.358	0.269
Min		0.590	0.623	0.456
Q1		0.795	0.771	0.579
Median		0.713	0.672	0.499
Q3		0.599	0.526	0.386
P95		0.509	0.444	0.326
Minkowski	p=1/8	0.820	0.806	0.610
	p=1/4	0.821	0.807	0.610
	p=1/2	0.822	0.808	0.612
	p=2	0.823	0.810	0.613
	p=4	0.820	0.809	0.611
	p=8	0.809	0.800	0.602
Percentile Pooling		0.823	0.810	0.613
Monotonic	p=1/8	0.823	0.810	0.613
	p=1/4	0.823	0.810	0.613
	p=1/2	0.823	0.810	0.613
	p=1	0.820	0.808	0.611
	p=2	0.811	0.801	0.603
	p=4	0.786	0.781	0.583
	p=8	0.738	0.738	0.544
FNS1		0.713	0.685	0.505
FNS2		0.726	0.698	0.516
FNS3		0.805	0.793	0.597
FNS4		0.804	0.792	0.596
FNS5		0.715	0.686	0.506
FNS6	$\lambda = 0$	0.806	0.797	0.601
FNS6	$\lambda = 0.1$	0.809	0.799	0.603
FNS6	$\lambda = 0.2$	0.811	0.800	0.604
FNS6	$\lambda = 0.3$	0.812	0.801	0.605
FNS6	$\lambda = 0.4$	0.812	0.801	0.605
FNS6	$\lambda = 0.5$	0.813	0.801	0.605
FNS6	$\lambda = 0.6$	0.813	0.801	0.604
FNS6	$\lambda = 0.7$	0.812	0.800	0.604
FNS6	$\lambda = 0.8$	0.812	0.800	0.604
FNS6	$\lambda = 0.9$	0.812	0.800	0.603
FNS6	$\lambda = 1$	0.811	0.799	0.602

TABLE III
SUMMARY OF RESULTS (SCIELAB AND 50 CM). SPEARMAN, PEARSON AND KENDALL CORRELATIONS AFTER NON-LINEAR REGRESSION.

Pooling Method		**CIDIQ Database (50 cm)**		
	Parameter	Pearson	Spearman	Kendall
Mean		0.657	0.646	0.473
Max		0.649	0.643	0.459
Min		0.214	0.217	0.153
Q1		0.454	0.394	0.286
Median		0.511	0.490	0.355
Q3		0.607	0.599	0.438
P95		0.718	0.706	0.521
Minkowski	p=1/8	0.454	0.481	0.347
	p=1/4	0.540	0.523	0.375
	p=1/2	0.586	0.577	0.417
	p=2	0.411	0.696	0.516
Percentile Pooling		0.658	0.647	0.474
Monotonic	p=1/8	0.678	0.666	0.488
	p=1/4	0.687	0.676	0.496
	p=1/2	0.696	0.680	0.501
	p=1	0.676	0.665	0.487
	p=2	0.625	0.626	0.451
	p=4	0.598	0.595	0.423
	p=8	0.610	0.605	0.428
FNS1		0.686	0.682	0.496
FNS2		0.692	0.687	0.500
FNS3		0.692	0.687	0.501
FNS4		0.672	0.662	0.487
FNS5		0.614	0.604	0.439
FNS6	$\lambda = 0$	0.693	0.683	0.504
FNS6	$\lambda = 0.1$	0.690	0.680	0.501
FNS6	$\lambda = 0.2$	0.686	0.676	0.498
FNS6	$\lambda = 0.3$	0.682	0.672	0.495
FNS6	$\lambda = 0.4$	0.677	0.666	0.490
FNS6	$\lambda = 0.5$	0.670	0.660	0.485
FNS6	$\lambda = 0.6$	0.662	0.651	0.477
FNS6	$\lambda = 0.7$	0.652	0.641	0.469
FNS6	$\lambda = 0.8$	0.638	0.626	0.457
FNS6	$\lambda = 0.9$	0.618	0.606	0.439
FNS6	$\lambda = 1$	0.587	0.575	0.415

metrics, and for the two viewing distances in CID:IQ.

The future work includes testing the different pooling methods on other publicly available databases, and also additional IQ metrics should be tested, since it has been reported that pooling is metric dependent [2].

REFERENCES

[1] Huib de Ridder. Minkowski-metrics as a combination rule for digital-image-coding impairments. In Bernice E. Rogowitz, editor, *Human Vision, Visual Processing, and Digital Display III*, volume 1666, pages 16 – 26. International Society for Optics and Photonics, SPIE, 1992.

TABLE IV
SUMMARY OF RESULTS (SCIELAB AND 100 CM). SPEARMAN, PEARSON AND KENDALL CORRELATIONS AFTER NON-LINEAR REGRESSION.

Pooling Method		**CIDIQ Database (100 cm)**		
	Parameter	Pearson	Spearman	Kendall
Mean		0.666	0.645	0.466
Max		0.621	0.673	0.483
Min		0.300	0.211	0.149
Q1		0.445	0.344	0.248
Median		0.530	0.458	0.328
Q3		0.638	0.595	0.431
P95		0.754	0.733	0.545
Minkowski	p=1/8	0.421	0.450	0.322
	p=1/4	0.543	0.497	0.353
	p=1/2	0.594	0.561	0.400
	p=2	0.724	0.714	0.528
Percentile Pooling		0.667	0.646	0.467
Monotonic	p=1/8	0.702	0.681	0.500
	p=1/4	0.719	0.699	0.515
	p=1/2	0.739	0.718	0.533
	p=1	0.736	0.724	0.539
	p=2	0.714	0.704	0.523
	p=4	0.676	0.668	0.486
	p=8	0.649	0.648	0.464
FNS1		0.708	0.706	0.518
FNS2		0.714	0.708	0.521
FNS3		0.713	0.709	0.521
FNS4		0.689	0.668	0.486
FNS5		0.625	0.596	0.427
FNS6	$\lambda = 0$	0.720	0.699	0.514
FNS6	$\lambda = 0.1$	0.715	0.694	0.510
FNS6	$\lambda = 0.2$	0.709	0.689	0.505
FNS6	$\lambda = 0.3$	0.703	0.682	0.499
FNS6	$\lambda = 0.4$	0.695	0.674	0.492
FNS6	$\lambda = 0.5$	0.686	0.665	0.483
FNS6	$\lambda = 0.6$	0.675	0.654	0.474
FNS6	$\lambda = 0.7$	0.661	0.640	0.462
FNS6	$\lambda = 0.8$	0.644	0.621	0.446
FNS6	$\lambda = 0.9$	0.621	0.594	0.424
FNS6	$\lambda = 1$	0.588	0.556	0.394

[2] Mingming Gong and Marius Pedersen. Spatial pooling for measuring color printing quality attributes. *Journal of Visual Communication and Image Representation*, 23:685–696, 07 2012.

[3] L. Cormack H.R. Sheikh, Z.Wang and A.C. Bovik. Live image quality assessment database release 2, http://live.ece.utexas.edu/research/quality, 2007.

[4] Dinesh Jayaraman, Anish Mittal, Anush K Moorthy, and Alan C Bovik. Objective quality assessment of multiply distorted images. In *2012 Conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR)*, pages 1693–1697. IEEE, 2012.

[5] Xinwei Liu, Marius Pedersen, and Jon Yngve Hardeberg. CID:IQ – a new image quality database. In Abderrahim Elmoataz, Olivier Lezoray, Fathallah Nouboud, and Driss Mammass, editors, *Image and Signal Processing*, pages 193–202, Cham, 2014. Springer International Publishing.

[6] Anush Krishna Moorthy and Alan Conrad Bovik. Visual importance pooling for image quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):193–201, 2009.

[7] Marius Pedersen. Evaluation of 60 full-reference image quality metrics on the CID:IQ. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 1588–1592. IEEE, 2015.

[8] Marius Pedersen and Jon Yngve Hardeberg. Full-reference image quality metrics: Classification and evaluation. *Foundations and Trends® in Computer Graphics and Vision*, 7(1):1–80, 2012.

[9] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo. Color image database tid2013: Peculiarities and preliminary results. In *European Workshop on Visual Information Processing (EUVIP)*, pages 106–111, 2013.

[10] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. 10(4):30–45, 2009.

[11] H.R. Sheikh, M.F. Sabir, and A.C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, 2006.

[12] Dogancan Temel and Ghassan AlRegib. A comparative study of quality and content-based spatial pooling strategies in image quality assessment. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 732–736, 2015.

[13] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[14] Zhou Wang and Xinli Shang. Spatial pooling strategies for perceptual image quality assessment. In *2006 International Conference on Image Processing*, pages 2945–2948, 2006.

[15] Cherinet G. Zewdie, Marius Pedersen, and Zhaohui Wang. A new pooling strategy for image quality metrics: Five number summary. In *2014 5th European Workshop on Visual Information Processing (EUVIP)*, pages 1–6, 2014.

[16] Xuemei Zhang and Brian A Wandell. A spatial extension of cielab for digital color-image reproduction. *Journal of the society for information display*, 5(1):61–63, 1997.