IMAGE SHARPNESS METRIC BASED ON MAXPOL CONVOLUTION KERNELS

Mahdi S. Hosseini and Konstantinos N. Plataniotis

The Edward S. Rogers Sr. Department of ECE, University of Toronto, ON M5S 3G4, Canada

ABSTRACT

We presents a no-reference (NR) image sharpness metric based on a visual sensitivity model. We propose that MaxPol convolution kernels are close approximation to this model and capable of extracting meaningful features for image sharpness assessment. Equipped by these kernels, we develop an efficient pipeline to evaluate the out-of-focus level of input images by decomposing the first and third order image differentials. The associated kernels are regulated in higher cutoff frequencies to balance out the information loss and noise sensitivity. We use high order central moments to exploit sharpness scores in wide range of frequency information. The experimental results outperform the state-of-the-art methods in accuracy and speed.

Index Terms— Visual sensitivity, MaxPol convolution kernels, No-reference image sharpness assessment

1. INTRODUCTION

The relative sensitivity of human visual cells are tuned with respect to different spatial frequencies, where it is shown that it is analogous to a power function magnitude in the frequency domain [1–3]. Provided by an image, the average response of visual cells at different frequencies is modeled by multiplying such a sensitivity response to the amplitude spectrum of the image for observation. Natural images usually follow a decay spectrum of $1/\omega$, hence causing a uniform response in the frequency domain observed by visual cells. When an image is out-of-focus, its amplitude spectrum will be steeper than $1/\omega$, and in conclusion, it attenuates the average response in the visual cells for higher frequencies. The implication of such visual perception model is intriguing to understand how human visual system (HVS) perform on subjective scoring for image quality assessment (IQA) [4, 5]. From engineering point of view, such model implies that HVS *"deconvolves"* visual content to grade the out-of-focus level of an image. In other words the power magnitude response of visual sensitivity boosts high frequencies for such assessment. In the context of signal processing this is equivalent to a convolution problem by filtering an image with impulse response of visual sensitivity function. A close approximation to this sensitivity response is made by finite impulse response (FIR) filters such as Laplacian or gradient operators that highly correlate with image sharpness scores made by human [6–9].

The numerical implementation of such FIR filters, however, are suboptimal and do not fully exploit the frequency range for approximation accuracy. In this paper we bridge this gap by introducing MaxPol filters [10, 11] as close approximation to visual sensitivity response, in which they are capable of approximating high order image differentials by introducing a balanced cutoff frequency to minimize the information loss and noise sensitivities. Equipped with MaxPol design, we tackle the problem of *no-reference image sharpness assessment (NRISA)* that has been dominating the field of imaging in the past few years [6–9, 12] to control the quality of acquired images in many applications for processing, storage and communication. Our proposed method for NRISA is of four-fold. First, a variational decomposition of gray image is made by MaxPol kernels of first and third order of derivatives. A cutoff frequency is regulated with the filters to balance the information loss and noise sensitivity. Second, a feature map is created by horizontal and vertical decomposition in $\ell_{1/2}$-norm space to promoted the sparsity of image edges. An adaptive threshold is introduced next to keep meaningful vector coefficients for processing. Whereas, it is followed by high-order central moment decomposition to extract informative features that are scattered in wide image frequencies. We examine the performance of the proposed method on four public subjectively-rated blur image database. Experiments demonstrates the unique advantage of the proposed method in terms of both accuracy and computational complexity.

2. CONVOLUTION FILTER DESIGN

In this section we design a new set of convolutional kernels that are scattered in different subbands of frequency domain. These filters are of high order of frequency polynomials, that are equivalent to high order differential operators in spatial domain. The response of this operator is determined by $\mathcal{F}(\partial^n/\partial x^n) = (i\omega)^n$ with $\mathcal{F}$ being the Fourier transform here. In particular, we are interested to approximate an ideal lowpass filter

$$H_{\omega_c}^n = \begin{cases} (i\omega)^n, & 0 \le \omega \le \omega_c \\ 0, & \text{else} \end{cases} \qquad (1)$$

where ω_c is the cutoff frequency. The Spectral response of such lowpass filter indicates that only the frequencies between $0 \le \omega \le \omega_c$ are amplified in magnitude of nth order polynomial, where the rest of the band is suppressed. We suspect that this is

ICIP 2018

a close approximation to visual sensitivity response modeled in [1–3]. For the purpose of numerical implementation we use Maximally Polynomial (MaxPol) library introduced in [10, 11] to generate variety of FIR kernels of different order of differentiation and cutoff frequencies. Examples of the filter spectra are shown in Figure 1 for different band designs. Here, the filter polynomial is 8 and the frequency resolution of subbands are $\pi/8$.
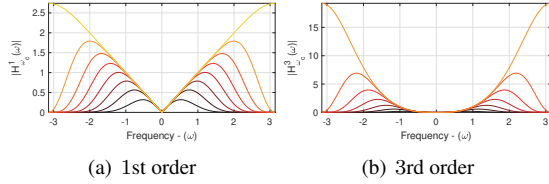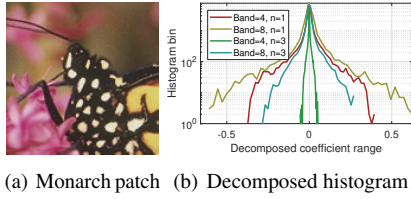

(a) 1st order   (b) 3rd order

Fig. 1. First and third order lowpass derivative filters designed by MaxPol library with different cutoff frequencies ω_c.

(a) Monarch patch (b) Decomposed histogram

(c) $\frac{\partial^1}{\partial y^\intercal}, \omega_c = \frac{\pi}{2}$ (d) $\frac{\partial^1}{\partial y^\intercal}, \omega_c = \pi$ (e) $\frac{\partial^3}{\partial y^3}, \omega_c = \frac{\pi}{2}$ (f) $\frac{\partial^3}{\partial y^3}, \omega_c = \pi$
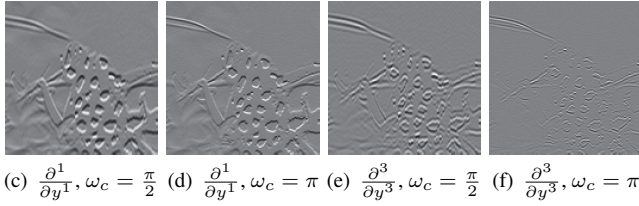
Fig. 2. Subband decomposition of Monarch image using 4th order of differentiation in two different bands. The image is decomposed in both horizontal and vertical directions.

The practival implication of these filters are they decompose different energy types in image edges. For instance, associated filter with response $H^3_{\frac{\pi}{2}}$ encodes edge types of third order moments containing harmonic frequencies in the range of $\omega \in [0, \frac{\pi}{2}]$. Figure 2 demonstrates examples of different order moments of decomposed Monarch image divided in different cutoff frequencies. As the band increases, high energy with sharp edges are encoded.

3. FOCUS QUALITY MEASURMENT

In this section we elaborate on the steps taken to build the focus quality metric from a discrete image in grayscale. The pipeline contains four main operations discussed in subsequent sections as follows.

MaxPol variational decomposition. We note the MaxPol lowpass FIR kernels by $d^n_{\omega_c}$ to approximate the transfer function in (1). The horizontal and vertical decomposition of an image $I \in \mathbb{R}^{N_1 \times N_2}$ using this kernel is obtained by

$$\nabla^n_{\omega_c} I = \left[I \star d^n_{\omega_c}, I \star d^n_{\omega_c}{}^T \right]^T. \tag{2}$$

The selection of cutoff band ω_c depends on the signal-to-noise-ratio (SNR) of the measurements. Higher cutoffs should be selected for high quality signals to avoid information loss. The tap-length polynomial of the FIR filters are chosen 8 here meaning that there are $\frac{\pi}{8}$ possible subband resolutions.

Feature map. Associated by the first and third order of derivatives $n \in \{1, 3\}$, the corresponding feature maps are calculated in $\ell_{\frac{1}{2}}$–norm space as follows

$$M^n_{\omega_c} = \left(|\nabla^n_{\omega_c} I(x)|^{\frac{1}{2}} + |\nabla^n_{\omega_c} I(y)|^{\frac{1}{2}} \right)^2. \tag{3}$$

The feature map $M^n_{\omega_c}$ encodes the edge significance in pertinent $\ell_{\frac{1}{2}}$–norm space. This promotes the sparsity of decomposed coefficients by weakening minor perturbations while preserving significant coefficients related to image edges.

Adaptive hard-thresholding. Similar to [13], we keep $|\Omega|$ dominant feature pixels from feature map $M^n_{\omega_c}$ to eliminate shallow coefficients that are less related to focus features

$$\overline{M}^n_{\omega_c} = sort_d(M^n_{\omega_c})_k, \ k \in \Omega \subset \{1, \ldots, N_1 N_2\} \tag{4}$$

where, $sort_d$ is a sorting operator in descending form and the notation $|\Omega|$ refers to cardinality of subset Ω. In addition, we adaptively select the number of remaining pixels $|\Omega|$ from the histogram distribution of absolute decomposed coefficients $|\nabla^n_{\omega_c} I|$ in (2) and map the approximated variance of distribution using a nonlinear activation function for adaptive selection of meaningful coefficients

$$p(\sigma) = \frac{1}{5}(1 - \tanh(50\sigma - 5)) + \frac{1}{25}. \tag{5}$$

Note that the approximated variance σ is normalized in scale of maximum amplitude of $|\nabla^n_{\omega_c} I|$. The image plot of this projection is shown in Figure 3. A lower value of σ is an indication of high sparse image in the derivative domain $\nabla^n_{\omega_c} I$ which we need to keep more pixel coefficients to exploit pertinent focus information. High σ is related to out-of-focus images where we keep less coefficients to avoid over fitting problem.
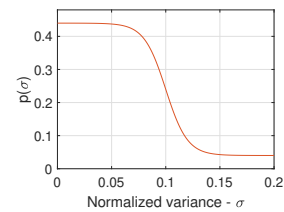

Fig. 3. Nonlinear mapping (sigmoid-like) function on normalized variance

Central moments information. The selected coefficients in (4) provide meaningful information from different layers

297

of image differentials (first and third orders in this study). Since we have deployed high accuracy MaxPol filters for such decomposition, the coefficients on edge locations contain high frequency components related to the evolution of image edge that contain high order moment information. To fully exploit information, it is required to measure high order moments from statistical distribution. To this end, we measure the mth central moment of $\overline{M}_{\omega_c}^n$ for feature extraction defined by

$$\mu_m = \mathbb{E}\left[(\overline{M}_{\omega_c}^n - \mu_0)^m\right] \qquad (6)$$

where, $\mathbb{E}$ is the expectation and $\mu_0 = \frac{1}{|\Omega|}\sum_{k\in\Omega}\overline{M}_{\omega_c}^n(k)$ is the average value of remaining features. The mth moment μ_m encodes the mth power of deviation of variables $\overline{M}_{\omega_c}^n$ from its mean μ_0. Separate measures are obtained for first and third order derivatives, where the final sharpness score is calculated by superimposing the feature moments in logarithmic scale

$$\text{Sharpness Score} = \log\mu_{m_1} + \log\mu_{m_3} \qquad (7)$$

where, μ_{m_1} and μ_{m_3} are the central moments evaluated on the first and third derivative orders in (6), respectively. The logarithm function converts the power measurements in (6) to linear scale. The tuning of the parameters m_1 and m_3 are individually analyzed to yield optimum correlation results with subjective scoring discussed in Section 4.

4. EXPERIMENTS

The empirical analysis of focus quality assessment here is performed on four blur image database, including, LIVE [5], CSIQ [14], TID2008 [15], and TID2013 [16], containing 145, 150, 100 and 125 Gaussian blurred images, respectively. The blurred images in LIVE and CSIQ are measured using difference mean opinion score (DMOS), and TID2008 and TID2013 are measured using mean opinion score (MOS). We use two criteria to evaluate the performance including Pearson linear correlation coefficient (PLCC) to measure prediction accuracy and Spearman rank order correlation (SRCC) to assess the prediction monotonicity, all defined in [5, 17]. To evaluate the latter performances, the objective scores are first mapped to the same scale of subjective scores using a nonlinear fitting model of five-parameter logistic regression function proposed in [5]. Throughout the experiments we used AMD FX-8370E 8-Core CPU 3.30 GHz in Windows station for processing.

Fine tuning of central moments. Of possible choices in (7), we need to select two combination of $\{m_1, m_2\}$ to yield the most correlation with subjective scores of database. Figure 4 demonstrates such analysis by selecting different combinations of order moments for all database. While we achieve optimum performance in LIVE using lower order moments for both derivatives, the rest of three database show better performance by increasing the order moments of first derivative decomposition. We found the parameters to be tuned

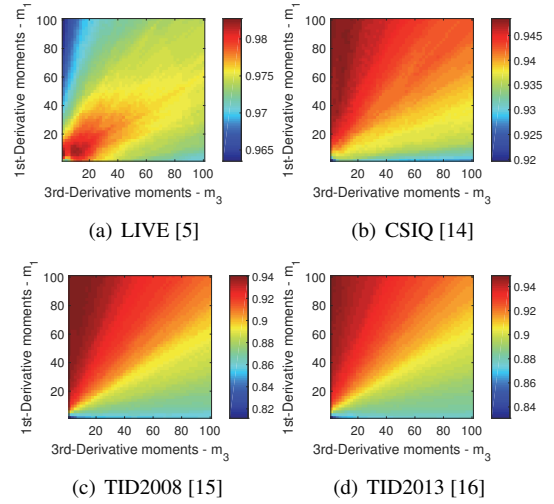at $m_1 = 72$ and $m_2 = 8$ to achieve optimum performance across all database.

(a) LIVE [5]   (b) CSIQ [14]

(c) TID2008 [15]   (d) TID2013 [16]

Fig. 4. SRCC performance evaluation of four database using different central moment on feature maps $\overline{M}_{\omega_c}^n$ created by first and third order of derivatives i.e. $n = \{1, 3\}$

Performance evaluation. We compare the MaxPol sharpness index against eight existing no-reference image sharpness metrics including S_3 [13], MLV [6], Kang's CNN [18], ARISM_C [19], GPC [20], SPARISH [7], RISE [9], and Yu's CNN [21]. The performance evaluation of all methods are studied in terms of correlation accuracies that are PLCC and SRCC, and computational complexity of the algorithms implemented for each method. This is measured by taking the average time for processing one image pixel in 8-bit gray level across the whole database, where low values indicates high speed performance. Table 1 summarizes the results obtained for above-mentioned methods. We highlight the top two rank methods in each criteria using bold ① light-green and ② dark-green colors, respectively. Furthermore, Table 1 computes the weighted averages of PLCC and SRCC over all database with respect to their size. We observe that the proposed MaxPol sharpness index score highly correlates with human visual perception and remarkably outperforms the other eight competing methods. In particular, we achieve 1% increase in overall PLCC and 1.73% increase in overall SRCC compared to the second top rank method among the competitors. Across four different database and two correlation measures, MaxPol achieves five times the first rank position and three times the second rank position compared to other existing methods. Aside the significance of correlation accuracy, MaxPol is intuitively fast for practical implementation, where it consumes minimal time for processing. For instance, it takes 0.45 second for MaxPol to create sharpness score for an image size of 1024×1024. Whereas, this consumes 6.74 second to process the same image size using the second best

298

Table 1. Focus score evaluation over four image database using different quality assessment methods. The performances are reported in terms of correlation accuracies PLCC and SRCC, and computational complexities in terms of time take to process one image pixel.

Method	Year	Measure	LIVE [5]	CSIQ [14]	TID2008 [15]	TID2013 [16]	Overall	CPU Time/Pixel (sec)
S_3 [13]	2012	PLCC	0.9434	0.9175	0.8555	0.8816	0.9042	5.2930×10^{-5}
		SRCC	0.9436	0.9058	0.8480	0.8609	0.8944	
MLV [6]	2014	PLCC	0.9590	0.9069	0.8584	0.8830	0.9063	3.1373×10^{-7}
		SRCC	0.9566	0.9246	0.8546	0.8785	0.9090	
Kang's CNN [18]	2014	PLCC	0.9625	0.7743	0.8803	0.9308	0.8848	*
		SRCC	0.9831	0.7806	0.8496	0.9215	0.8842	
ARISM$_C$ [19]	2015	PLCC	0.9590	0.9481	0.8544	0.8979	0.9211	7.2782×10^{-5}
		SRCC	0.9561	0.9314	0.8681	0.9015	0.9189	
GPC [20]	2015	PLCC	0.9242	0.9018	0.8684	0.8665	0.8931	3.2723×10^{-7}
		SRCC	0.8369	0.8641	0.8729	0.8668	0.8589	
SPARISH [7]	2016	PLCC	0.9595	0.9380	0.8900	0.9020	0.9261	1.4906×10^{-5}
		SRCC	0.9593	0.9139	0.8836	0.8940	0.9159	
RISE [9]	2017	PLCC	0.9620	0.9463	0.9289	0.9419	0.9463	6.4270×10^{-6}
		SRCC	0.9493	0.9279	0.9218	0.9338	0.9341	
Yu's CNN [21]	2017	PLCC	0.9730	0.9416	0.9374	0.9221	0.9449	*
		SRCC	0.9646	0.9253	0.9189	0.9135	0.9322	
MaxPol	2018	PLCC	0.9735	0.9657	0.9359	0.9412	0.9563	4.2911×10^{-7}
		SRCC	0.9688	0.9481	0.9394	0.9448	0.9514	

∗: the algorithm is implemented in GPU card with high speed and comparison here does not apply. Please refer to corresponding reference for more information.
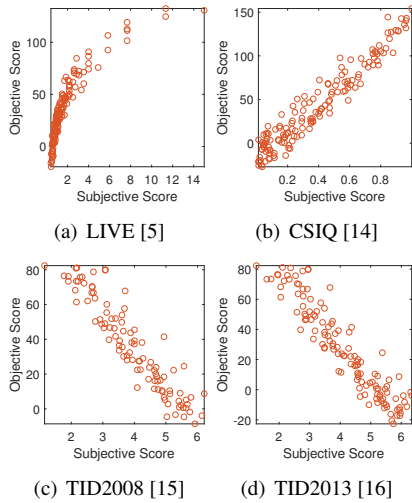
method (RISE) for scoring.

(a) LIVE [5] (b) CSIQ [14]

(c) TID2008 [15] (d) TID2013 [16]

Fig. 5. Scatter plots between subjective and objective scores. The objective scores are raw with no mapping.

The objective scores obtained from MaxPol method are plotted in Figure 5 where there is no nonlinear functional mapping applied to make them linear. In fact, the proposed sharpness score is corresponding linearly with respect to subjective scores in three database of CSIQ, TID2008, and TID2013. Such linear response is an indication of complying with the reality of subjective score based on human visual assessment.

Noise sensitivity analysis. Our final analysis here is carried out by considering independent and identically distributed (i.i.d) Gaussian noise on input image for evaluation. This type of noise is also considered as real noise application in many natural imaging problems. Our goal here is to analyze the sensitivity of MaxPol convolution kernels with respect such

noise artifacts. One of the main merits of the MaxPol kernels is noise-robust approximation of derivative features due to ability of setting cutoff frequency for numerical differentiation. The Monte-Carlo simulation is done here for 10 average simulations on generating random perturbations with different standard deviation magnitudes $\sigma \in \{0.005, 0.01, \dots, 0.05\}$. These perturbations are added to blurred images of selected database prior to sharpness score evaluation. Figure 6 demonstrates this simulation over all database to evaluate the PLCC by means of different cutoff frequencies for MaxPol design. As it shown, lower cutoffs are required to mitigate the correlation loss over high noise perturbations.
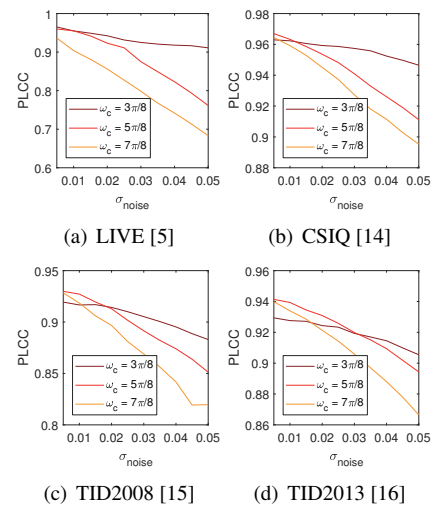

(a) LIVE [5] (b) CSIQ [14]

(c) TID2008 [15] (d) TID2013 [16]

Fig. 6. PLCC performance on four database simulated in different white Gaussian noise levels σ_{noise}

5. REFERENCES

[1] David J Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Josa a*, vol. 4, no. 12, pp. 2379–2394, 1987.

[2] Nuala Brady and David J Field, "What's constant in contrast constancy? the effects of scaling on the perceived contrast of bandpass patterns," *Vision research*, vol. 35, no. 6, pp. 739–756, 1995.

[3] David J Field and Nuala Brady, "Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes," *Vision research*, vol. 37, no. 23, pp. 3367–3383, 1997.

[4] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[5] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on image processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[6] Khosro Bahrami and Alex C Kot, "A fast approach for no-reference image sharpness assessment based on maximum local variation," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 751–755, 2014.

[7] Leida Li, Dong Wu, Jinjian Wu, Haoliang Li, Weisi Lin, and Alex C Kot, "Image sharpness assessment by sparse representation," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1085–1097, 2016.

[8] Leida Li, Weisi Lin, Xuesong Wang, Gaobo Yang, Khosro Bahrami, and Alex C Kot, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE transactions on cybernetics*, vol. 46, no. 1, pp. 39–50, 2016.

[9] Leida Li, Wenhan Xia, Weisi Lin, Yuming Fang, and Shiqi Wang, "No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features," *IEEE Transactions on Multimedia*, vol. 19, no. 5, pp. 1030–1040, 2017.

[10] M. S. Hosseini and K. N. Plataniotis, "Derivative kernels: Numerics and applications," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4596–4611, Oct 2017.

[11] Mahdi S. Hosseini and Konstantinos N. Plataniotis, "Finite differences in forward and inverse imaging problems: Maxpol design," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1963–1996, 2017.

[12] Rania Hassen, Zhou Wang, and Magdy MA Salama, "Image sharpness assessment based on local phase coherence," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2798–2810, 2013.

[13] Cuong T Vu, Thien D Phan, and Damon M Chandler, "S_3: A spectral and spatial measure of local perceived sharpness in natural images," *IEEE transactions on image processing*, vol. 21, no. 3, pp. 934–945, 2012.

[14] Eric Cooper Larson and Damon Michael Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011006, 2010.

[15] 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," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.

[16] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., "Image database tid2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.

[17] Video Quality Experts Group et al., "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii," 2003.

[18] Le Kang, Peng Ye, Yi Li, and David Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1733–1740.

[19] Ke Gu, Guangtao Zhai, Weisi Lin, Xiaokang Yang, and Wenjun Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3218–3231, 2015.

[20] Arthur Leclaire and Lionel Moisan, "No-reference image quality assessment and blind deblurring with sharpness metrics exploiting fourier phase information," *Journal of Mathematical Imaging and Vision*, vol. 52, no. 1, pp. 145–172, 2015.

[21] Shaode Yu, Shibin Wu, Lei Wang, Fan Jiang, Yaoqin Xie, and Leida Li, "A shallow convolutional neural network for blind image sharpness assessment," *PloS one*, vol. 12, no. 5, pp. e0176632, 2017.

300