

SI-FID: Only One Objective Indicator for Evaluating Stitched Images

Xinrui Zhang^{1†}, Shengwei Guo^{1†}, Guobing Sun^{1*}

¹College of Electronic Engineering, Heilongjiang University, Haxi Street, Harbin, 150006, Heilongjiang, China.

*Corresponding author(s). E-mail(s): sunguobing@hlju.edu.cn;

†These authors contributed equally to this work.

Abstract

Image quality evaluation accurately is vital in developing image stitching algorithms as it directly reflects the algorithm's progress. However, commonly used objective indicators always produce inconsistent and even conflicting results with subjective indicators. To enhance the consistency between objective and subjective evaluations, this paper introduces a novel indicator—the Fréchet Distance for Stitched Images (SI-FID). To be specific, our training network employs the contrastive learning architecture overall. We employ data augmentation approaches that serve as noise to distort images in the training set. Both the initial and distorted training sets are then input into the pre-training model for fine-tuning. We then evaluate the altered FID after introducing interference to the test set and examine if the noise can improve the consistency between objective and subjective evaluation results. The rank correlation coefficient is utilized to measure the consistency. SI-FID is an altered FID that generates the highest rank correlation coefficient under the effect of a certain noise. The experimental results demonstrate that the rank correlation coefficient obtained by SI-FID is at least 25% higher than other objective indicators, which means achieving evaluation results closer to human subjective evaluation.

Keywords: quality of stitched image, consistency of subjective and objective indicator, contrastive learning, noise, fine-tuning

1 Introduction

Image quality evaluation is an essential process in computer vision that entails comparing generated images with their source images based on several factors such as clarity, contrast, resolution, etc. It is widely utilized in image generation fields like image stitching and image super-resolution to test the advancement of the algorithm.

There are two primary methods to evaluate the quality of an image: subjective and objective evaluation. The subjective evaluation relies on the personal perception of the individual who evaluates the image. This method uses specific rules and an experimental environment to judge the image quality by giving scores or levels. The mean opinion score (MOS) is the most commonly used method for subjective evaluation, which judges the image quality by normalizing the scores provided by the evaluators. And the International Telecommunication Union (ITU-T) provides guidelines for image testing, including observation distances and experimental environments for subjective evaluations[1]. The objective evaluation utilizes specific formulas or models designed to compute a score representing the image quality. The assessment items for objective indicators include the color, brightness, depth information, and texture characteristics of the image.

Objective evaluation may vary slightly due to differences in image generation algorithms. For stitched images, there are two types of objective evaluations: full reference evaluation (FR) and no reference evaluation (NR)[2]. The former requires a reference image to compare similarity with the stitched image. The latter aims to simulate the subjective evaluation process mentioned above to evaluate the image directly.

Stitched images are different from other generated images in the sense that they are formed by combining at least two images with overlapping information. As a result, most stitched images have misalignment or ghosting. The misalignment or ghosting phenomenon is due to (i) differences in angle, brightness, or color between inputs and (ii) geometric transformations experienced during the stitching process, and the defects are inevitable but can be clearly perceived by humans. Therefore, the subjective evaluation method is usually combined with objective evaluation for a comprehensive evaluation to ensure the evaluation results are more convincing. Objective indicators with higher consistency between quality and subjective evaluation results are more suitable as evaluation indicators for stitched images. The evaluation of the quality of stitched images involves full reference indicators such as mean square error (MSE)[3], peak signal-to-noise ratio (PSNR)[4], and structural similarity (SSIM)[5]. And the no reference indicators such as blind/referenceless image spatial quality evaluator (BRISQUE)[6] and natural image quality evaluator (NIQE)[7].

However, we found it challenging to use objective evaluation measures to capture the misalignments or ghosts in stitched images. The results of objective evaluations may not always match up with subjective evaluations. And in some cases, it may even lead to conflicting conclusions. As depicted in Figure 1, is the comparison of the quality evaluation results of a group of stitched images between subjective and objective evaluations. The leftmost column represents subjective evaluation and the above objective indicators. The numbers "1" to "4" represent the ranking of human and objective indicators in descending order of the quality of the stitched image, with

1 being the highest image quality and 4 being the lowest. It is not difficult to find a significant difference between the objective and subjective evaluation results, which makes the quality evaluation conclusion contradictory and confusing.

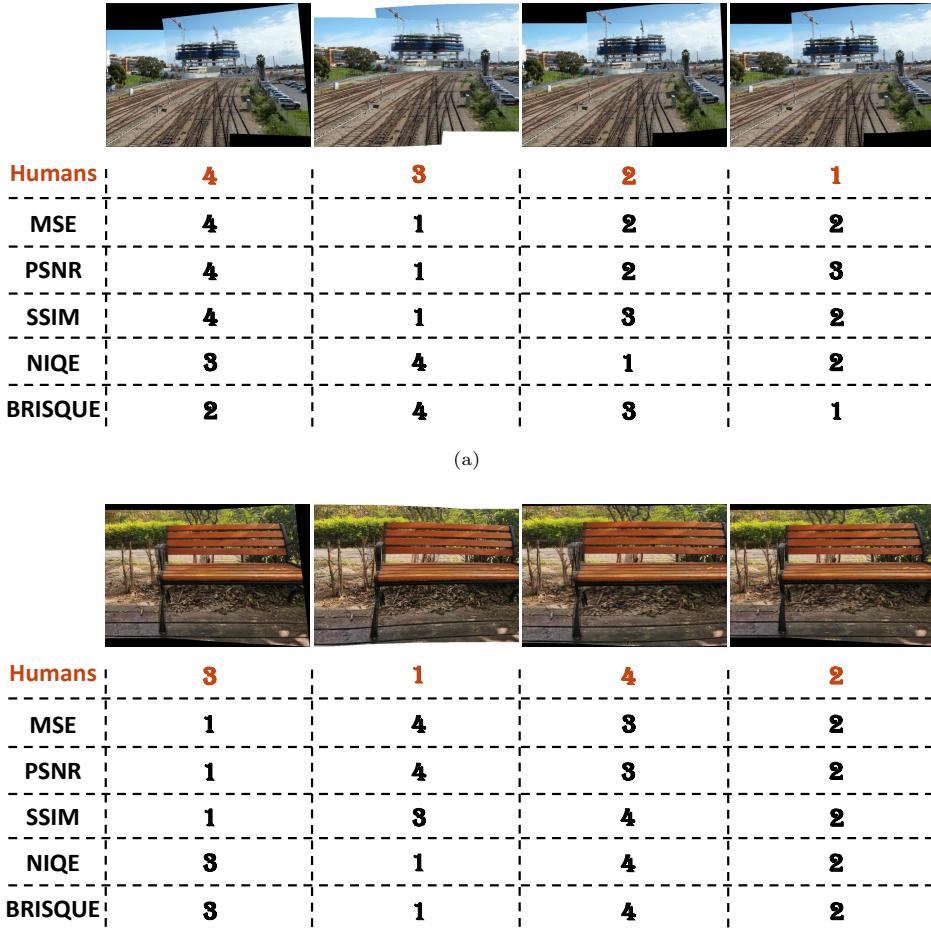


Fig. 1 The inconsistency between subjective and objective evaluations, the source images for generating stitched images are all from the test set.

Therefore, we design an indicator called SI-FID based on the Fréchet distance (FID)[8] to tackle the issue of objective indicators failing to account for misalignments and ghosts. The analysis has revealed that the quality evaluation process of FID indicates the similarity to contrastive learning[9], both of them holding the belief that images with similar features are likely to be similar to each other. Consequently, we propose defining quality evaluation for stitched images as a metric learning problem[10], where the distance between the respective image features determines the similarity. To improve the ability of FID to detect misalignments and ghosts, we involve

artificial noise through data augmentation to distort the images from the training set, which are then used in feeding the pre-training model InceptionV3 to fine-tune. SI-FID is a fine-tuning result interfered by noise. The fine-tuning schematic is shown in Figure 2.

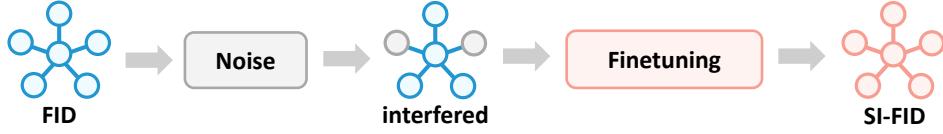


Fig. 2 Fine-tuning schematic.

This paper employs average rank correlation coefficients to measure the consistency between objective and subjective evaluation results. A higher average rank correlation coefficient for an objective indicator indicates that its evaluation results are closer to the subjective evaluation. The results show that SI-FID has a higher average rank correlation coefficient than other classical objective indicators such as PSNR, SSIM, etc.

2 Previous work

This section has provided a review of the previous studies on objective indicators. All indicators in the following content can be used to evaluate stitched image quality.

According to the different focuses of image quality evaluation, objective indicators can be used to evaluate the quality from four aspects: image structure, mutual information, feature distance, and visual fidelity.

Via the structure of figures, the average gradient (AG) and spatial frequency (SF) respectively examine the average gradient and grayscale change rate to reflect image clarity. Kalyan et al. proposed the mean squared error (MSE)[3], which is used to measure the difference in pixel grayscale values between images. The peak signal-to-noise ratio (PSNR)[4] proposed by Huynh-Thu et al. computes the ratio between the maximum value of image pixels and the MSE. In addition, Wang et al. proposed the structural similarity (SSIM)[5], which comprehensively judges the quality of images through brightness, contrast, and structure. On this basis, Zhang et al. proposed the feature similarity (FSIM)[11], which measures local similarity through the phase consistency and gradient features and then computes the similarity score via weighting the local similarity.

The mutual information (MI)[12] is a metric used to assess the quality of merged images. It measures the amount of information present in the input images by considering their information entropy and joint information entropy in the merged image. Following this, Haghigat et al. proposed the feature mutual information (FMI)[13], a metric that evaluates the information present in the merged image. However, it represents this information using image gradients or edges. Additionally, Piella et al. proposed Qabf[14], which uses local measures to estimate how well the critical information from the input images is present in the merged images.

Haghigat et al. proposed the concept of visual fidelity (VIFF)[15], which focuses more on the perception of brightness, contrast, color, and other aspects of stitched images by the Human Vision System (HVS) compared to the mentioned indicators. Zhang et al. developed a perceptual loss model called learned perceptual image patch similarity (LPIPS)[16] that mimics human visual perception to improve the consistency between objective and subjective evaluation results. Furthermore, more non reference indicators will imitate HVS for quality evaluation. Moorthy et al. proposed the blind image quality index (BIQI)[17], which uses the wavelet coefficients of the input image to statistically analyze image features. Mittal et al. designed the natural image quality evaluator (NIQE)[7], which is based on the construction of a ‘quality aware’ collection of statistical features based on the Natural Scene Statistical Model (NSS) to statistics the image features of natural images and distorted images. Meanwhile, Mittal et al. designed the blind/referenceless image spatial quality evaluator (BRISQUE)[6]. It quantifies the possible loss of naturalness in images due to distortions using scene statistics of locally normalized luminance coefficients, leading to a holistic quality measure. Venkatanath et al. proposed the perception-based image quality evaluator (PIQUE)[18], an evaluator that scores each image block using “distortion perception”, without any statistical learning.

The similarity of images can be measured using histograms or cosine similarity. The former compares the histograms, while the latter computes the cosine distance between vectors. Heusel et al. proposed the Fréchet distance (FID)[8], assuming the image features follow a Gaussian distribution and then calculating the Fréchet distance between two Gaussian distributions.

The above are the universal indicators of image quality evaluation, and a relatively single focus on quality evaluation leads to the inability to capture misalignment or ghosting in stitched images. Therefore, it is challenging to draw evaluation conclusions that match subjective evaluations. In recent studies, numerous methods have been proposed to align ghosts at the seams in stitched images accurately. Pavan et al. utilized a Gaussian mixture model to create a stitched image quality evaluator (SIQE)[19]. At the same time, Zhou et al. introduced a method to evaluate the quality of dual point-feature matching[20]. To capture distortions, Yang et al. combined perceptual geometric error measurement with local structural guidance measurement[21], and Ling et al. employed convolutional sparse encoding and convolutional filters to orientate distortions in the target image[22].

The above objective indicators designed for the characteristics of stitched images will undoubtedly yield more accurate evaluations. Prompted by this, the paper adopts the contrastive learning architecture to design an objective indicator called SI-FID for stitched images, which aims to measure the similarity between images through the feature distance between image features. The closer the distance stands for the higher the similarity. To ensure accurate detection of misalignment and ghosting, the paper injects artificial noise (distorted images by data augmentation) into the training process, which enhances the ability of SI-FID to perceive distorted areas. This approach is expected to yield more accurate evaluations of stitched images.

This paper aims to continuously explore how the evaluation of new indicators generated by adding noise will change compared to the original FID and seek to find

a certain type of noise, making the objective evaluation compliant with the subjective evaluation.

3 Fine-tuning Strategy

Adding noise into the pre-training has proven to yield favorable fine-tuning results. A case is a NoisyTune model proposed by Wu et al.[23], which helps better fine-tune PLMs on downstream tasks by adding some noise to the parameters of PLMs before finetuning. Similarly, the paper applies the FID indicator as a pre-training model and utilizes a range of data augmentation approaches to introduce distortions into the initial training set, making the initial and distorted training sets participate in fine-tuning.

FID belongs to the FR indicator that assesses the quality by computing the distance between the feature vectors of the source image and the stitched image. Specifically, the FID utilizes the InceptionV3[24] network model to extract the image features and compute the mean and variance, assuming that image features follow the Gaussian maximum entropy distribution. Then, the similarity between the binary Gaussian function representing the original image and the stitched image by the Fréchet distance. As shown in Figure 3, the image quality evaluation process for FID indicators is presented.

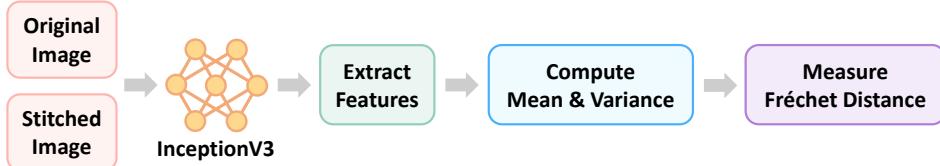


Fig. 3 Image quality assessment by FID indicator.

The paper has fine-tuned the FID, the network architecture is inspired by contrastive learning[25], as depicted in Figure 4. The architecture accepts the initial image x and an augmented image \tilde{x} as input. These two images are processed by an encoder network f , which includes the feature extraction layer of InceptionV3 and shares weights between the images. Denoting the two output image features as $F \triangleq f(x)$ and $\tilde{F} \triangleq f(\tilde{x})$. The cosine similarity function has been utilized as the loss function. The function is expressed through the following formula:

$$L = \frac{1}{2}D(F, \tilde{F}) = \frac{1}{2} \left[\frac{F}{\|F\|_2} \cdot \frac{\tilde{F}}{\|\tilde{F}\|_2} \right] \quad (1)$$

where $\|\cdot\|_2$ is ℓ_2 - norm.

Five data augmentation approaches, such as GaussianBlur, RandomResizedCrop, RandomGrayscale, ColorJitter, and RandomHorizontalFlip, are applied initially. But we expand the approaches to fourteen types by altering

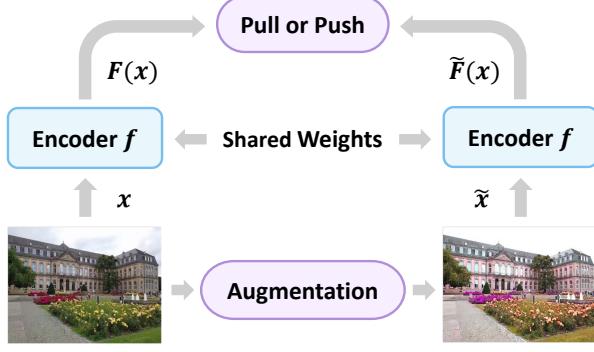


Fig. 4 Training for fine-tuning.

parameters, including GaussianBlur(3), GaussianBlur(13), GaussianBlur(39), RandomHorizontalFlip(0.5), RandomHorizontalFlip(0.8), RandomGrayscale(0.8), ColorJitter(brightness=0.3, hue=0.1), ColorJitter(brightness=0.5, hue=0.3), RandomResizedCrop(39), RandomResizedCrop(50), RandomResizedCrop(100), RandomResizedCrop(120), RandomResizedCrop(150), and RandomResizedCrop(190). Figure 5 shows the images of the training set under data augmentation.



Fig. 5 The initial and distorted training set. Figure 5(a) shows the initial training image. The top line in Figure 5(b) shows the color transformation through data augmentation, and the bottom line shows the geometric transformation through data augmentation.

This paper assumes that data augmentation does not substantially modify similar image features. When two images have a higher similarity, their features are positioned in closer proximity. We utilize the Stochastic Gradient Descent (SGD) algorithm to optimize the parameters. Its stochastic nature makes the training more robust by randomly selecting a small batch of samples in each iteration to compute the gradient of the loss function and update the parameters.

The epoch is set to 100 for each iteration and collocating SGD with a momentum of 0.9 for the parameters optimization. After each iteration, 100 altered FID indicators are produced. The batch size is 32. The learning rate is initialized to 0.01. By manually adjusting the learning rate, the training is repeated 312 times.

4 Experiments and Analysis

4.1 Dataset

The paper has three types of image sets: the training set, testing set, and subjective evaluation set (subjective critics obtain subjective evaluation results by observing them).

The training set is derived from the DIR-D dataset[26], which has a wide range of irregular scenes and rich detail features. Every image in the training set has a resolution of 512×384 . We selected 519 images from the DIR-D dataset as the initial training set and subsequently generated 7266 distorted images by data augmentation. Our training set totally contains 7785 images.

This paper has two test sets, each including source images, transformed images, and stitched images with the same scene information and resolution as the transformed images. The transformed and stitched images are generated by four algorithms: APAP[27], SPHP[28], ANAP[29], and SPW[30]. Concretely speaking, the transformed image is the source image after geometric transformation, and the stitched image is the final result of the above algorithms.

Test set I contains 20 sets of source images, 80 sets of transformed images, and 80 sets of stitched images, all of which are sourced from some classic image stitching algorithms. Test set II contains 20 sets of source images, 80 sets of transformed images, and 80 sets of stitched images, among which the source images are captured by ourselves. However, it should be noted that the quality evaluation is based on the similarity between the stitched and transformed images. The source image is only employed to generate transformed and stitched images.

The subjective evaluation set consists of 160 stitched images from the above test set.

4.2 Performance Assessment

4.2.1 Subjective Evaluation

This paper considers subjective evaluation results as significant factors in measuring whether objective indicators are advanced. The subjective critics are composed of 14 computer vision researchers, who are asked to view the stitched images from the same environment and provide the subjective scores representing image quality by observing the color, sharpness, and whether the misalignments or ghosts exist. The subjective score is limited from 0 to 100 continuous linear closed intervals.

However, the subjective scores have to be normalized to make them in the same dimension as the objective scores in the following text, which can ensure comparability between the scores. The scores after normalization are both limited from 0 to 1 continuous linear closed intervals. Each stitched image has been assessed 14 times so that the subjective scores have to be averaged to obtain only one score that illustrates its quality. The following formulas show the process of normalization and averaging for a certain stitched image:

$$x'_i = \frac{x_i - \mu}{\delta} \quad (2)$$

where $x_i (i = 1, 2, \dots, 14)$ represents the quality score given by subjective critics, the average of the 14 scores is represented by μ , and δ is the corresponding standard deviation,

$$S = \frac{\sum_{i=1}^{14} x'_i}{n} \quad (3)$$

where $n = 14$ represents the number of subjective critics, $x'_i (i = 1, 2, \dots, 14)$ represents the quality score after normalization, and S is the subjective score.

4.2.2 Which Noise Brings Benefits?

This paper employs the Pearson Rank Correlation Coefficient (PCC)[31] and the Spearman Rank Correlation Coefficient (SROCC)[32] to measure the correlation between objective score and subjective score, which is computed as follows:

$$r_p = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} \quad (4)$$

where X and Y stand for the objective score and subjective score respectively,

$$r_s = 1 - \frac{6\sum d_i^2}{X(X^2 - 1)} \quad (5)$$

where X also stands for the objective score, d stands for the rank value of X and the subjective score. The rank correlation coefficients fall within the closed interval -1 to 1, with a score closer to 1 denoting a higher degree of consistency between the objective and subjective evaluation results.

Next, this paper assesses the effectiveness of both the initial and altered FID on the test set. Specifically, the new FID and the original FID need to evaluate the quality of images within the test set I and generate the corresponding objective score. Subsequently, the new FID will be selected as our new indicator that has the highest consistency between the objective score and the subjective score.

To insight into the relationship between the new FID and the original FID intuitively, we have depicted the PCC and SROCC curves of the initial FID and the altered FID created by noise in Figure 6. The horizontal axis "0" represents the original FID, which means the iteration is zero. The horizontal axis "1" to "100" represents different iterations, that is, the new FID generated with the same noise but different iterations. For instance, the horizontal axis "10" indicates the new FID that is generated with 10 iterations. The vertical axis is PCC or SROCC.

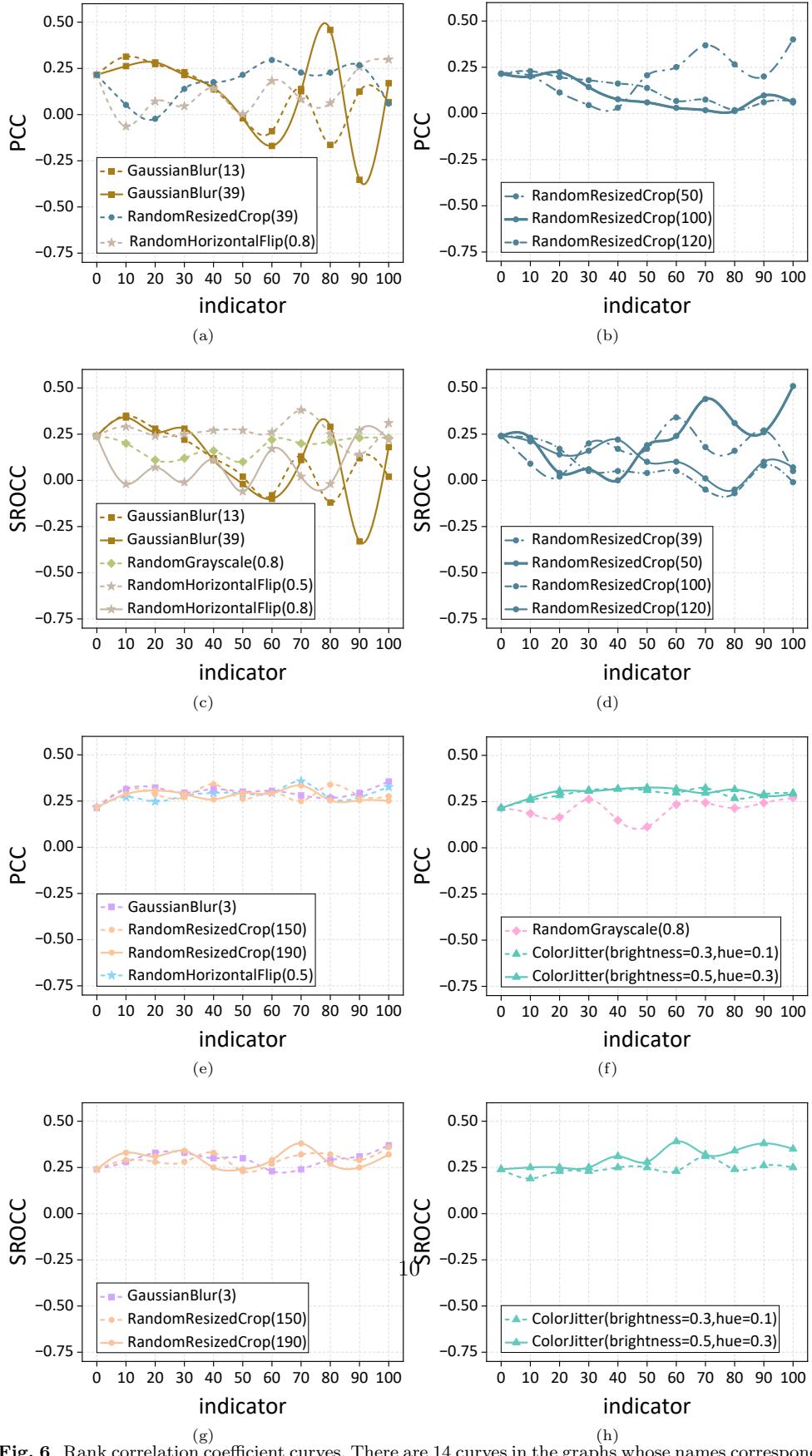


Fig. 6 Rank correlation coefficient curves. There are 14 curves in the graphs whose names correspond to the data augmentation approaches employed as noise.

According to Figure 6(a) to 6(d), which show the curves depicted through the "negative noise" exhibiting decreasing trends and considerable amplitudes. Conversely, Figure 6(e) to 6(h) show the curves depicted through the "positive noise", which meets the above characteristics. Therefore, Gaussian Blur (3), ColorJitter (brightness=0.5, hue=0.3), and RandomResizedCrop (150) regard as positive noise.

Next, this paper compares the curves of the altered FID after the "positive noise" once more, the comparison result is shown in Figure 7.

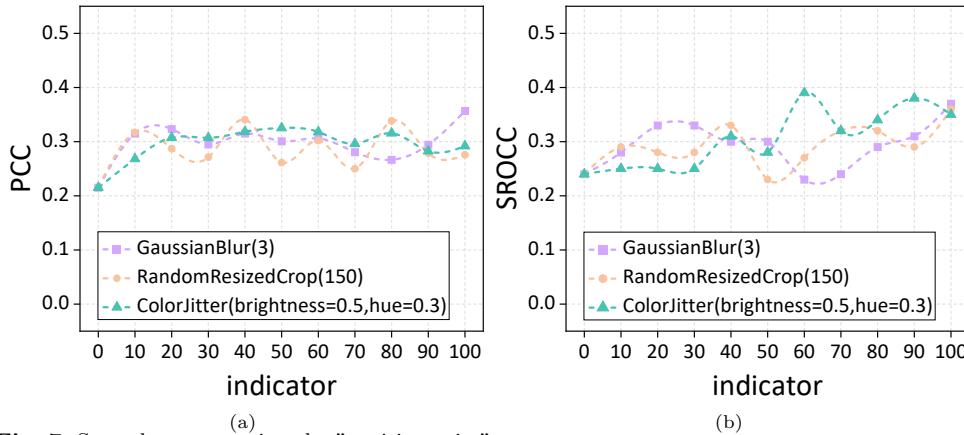


Fig. 7 Secondary comparison by "positive noise".

It is evident from Figure 7 that the curve called "ColorJitter (brightness=0.5, hue=0.3)" retains the smoothest and monotonic increase trend, which means the noise "ColorJitter (brightness=0.5, hue=0.3)" has the highest robustness. Our SI-FID is thus the altered FID with the highest PCC and SROCC subjected to the disturbance of ColorJitter (brightness=0.5, hue=0.3).

4.2.3 Comparison with State-of-the-art

In this section, a comparison is made between the capacity of the SI-FID and traditional objective indicators. All the indicators necessitate a quality evaluation of the test set owing to measuring the consistency between objective and subjective evaluation results. Note that the correlation value is closer to 1; the objective evaluation result is more approximate to human subjective judgments.

The classic objective indicators include FR indicators such as MSE, PSNR, SSIM, FSIM, LPIPS, initial FID, and no-referenced indicators such as NIQE and BRISQUE. As for FR indicators, MSE, PSNR, SSIM, and FSIM are traditional indicators that evaluate images' brightness, texture details, and clarity with formulas. LPIPS and initial FID are defined as Network-based objective indicators, which one or more network models obtain. The indicators have learned how to evaluate image quality. Remarkably, the evaluation rule for MSE, LPIPS, initial FID, NIQE, BRISQUE, and our SI-FID is "the lower the score, the higher similarity". Hence, we have to ensure that the objective scores of all indicators are located in the same dimension.

It is crucial to evaluate the image quality in both test sets I and II to establish the credibility of the result and the generalization of the SI-FID. In Figure 8, we present the average rank correlation coefficient of all indicators in test sets I and II. A higher rank correlation coefficient demonstrates a higher degree of consistency between the objective and subjective evaluation results.

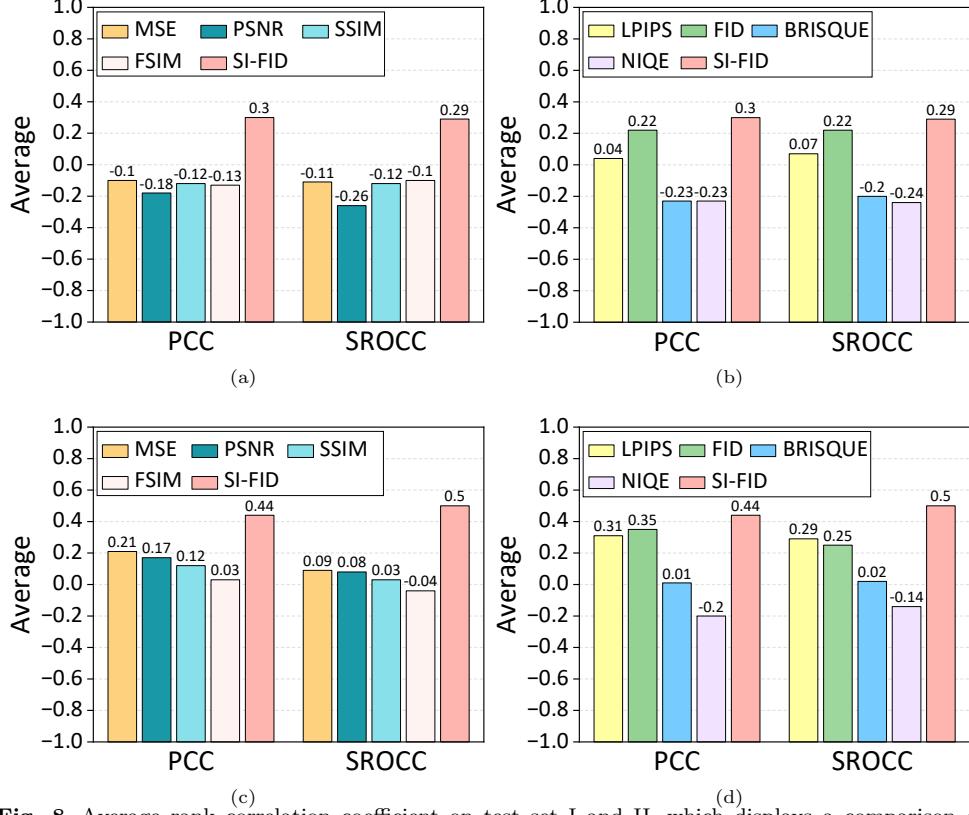


Fig. 8 Average rank correlation coefficient on test set I and II, which displays a comparison of traditional objective indicators, network-based objective indicators, NR objective indicators, and SI-FID. The top line is the result of test set I, while the bottom line is of test set II.

We can find in Figure 8 that SI-FID generates the highest value of PCC and SROCC, which means it performs the highest consistency with the subjective result. This finding further proves our conclusion in section 4.2.2 that ColorJitter (brightness=0.5, hue=0.3) acts as a "positive noise".

Moreover, we test the variance of the rank correlation coefficient of all indicators in to observe their stability during the quality evaluation. The lower the variance, the more robust the evaluation process of the objective indicator becomes.

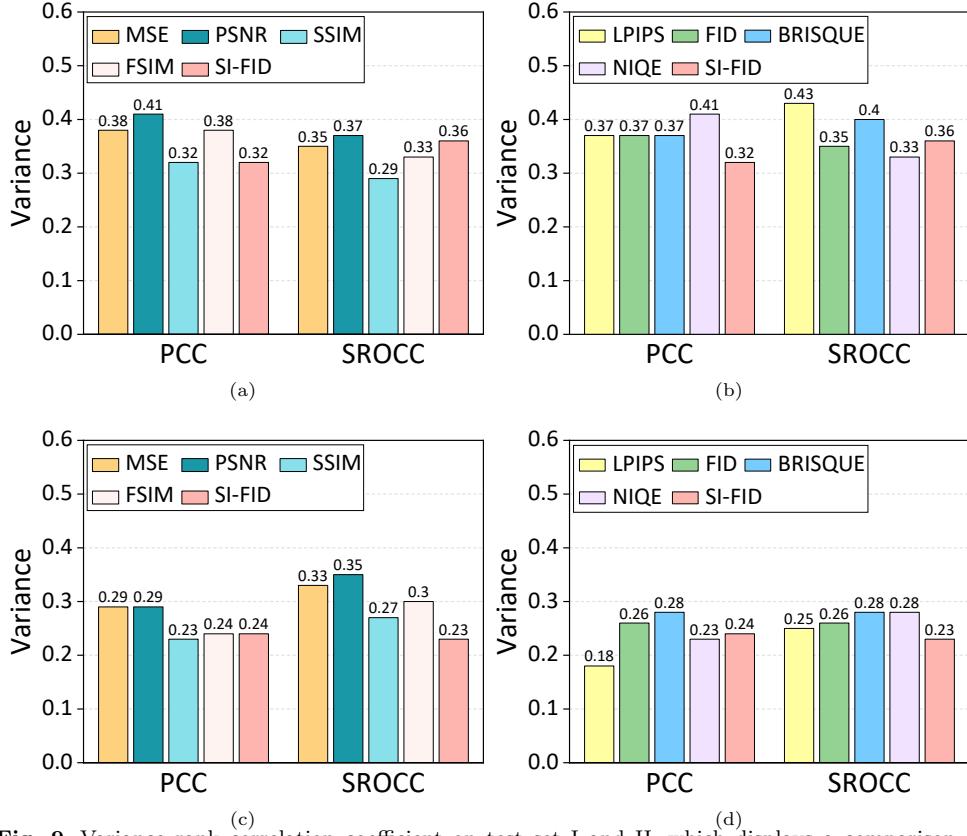


Fig. 9 Variance rank correlation coefficient on test set I and II, which displays a comparison of traditional objective indicators, network-based objective indicators, NR objective indicators, and SI-FID. The top line is the result of test set I, while the bottom line is of test set II.

From Figure 9, the PCC variance of SI-FID is the smallest in test set I, and the SROCC variance also performs well. The PCC and SROCC variance of SI-FID in test set II rank top. In general, SI-FID surpasses other conventional objective indicators in its consistency and resilience, meaning it is a more appropriate indicator for the quality evaluation of stitched images.

5 Conclusion

This paper investigates whether adding noise can enhance the consistency between objective and subjective evaluation results. We take a variety of data augmentation approaches that play as noise to participate in the fine-tuning of the initial indicator FID, then assess the rank correlation coefficient between the altered FID and the initial FID. The findings show that the inclusion of specific noise, such as Gaussian Blur (3), ColorJitter (brightness=0.5, hue=0.3), and RandomResizedCrop (150), have a positive effect on the fine-tuning process that can improve consistency between objective and subjective results on the fine-tuning process. Upon comparison, we determine

that ColorJitter (brightness=0.5, hue=0.3) is the most effective noise for fine-tuning. Hence, the new indicator SI-FID is the altered FID that demonstrates the highest rank correlation coefficient under the effect of ColorJitter (brightness=0.5, hue=0.3).

Furthermore, we compare the average rank correlation coefficient and the variance of the rank correlation coefficient between SI-FID and classical objective indicators in test sets I and II. The experimental results indicate that SI-FID performs best. Therefore, the evaluation results of SI-FID stitched images are closer to the subjective evaluation results of humans, and the robustness of the evaluation results is also guaranteed.

References

- [1] Hontsch, I., Karam, L.J.: Adaptive image coding with perceptual distortion control. *IEEE Transactions on Image Processing* **11**(3), 213–222 (2002) <https://doi.org/10.1109/83.988955>
- [2] Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing* **15**(11), 3440–3451 (2006) <https://doi.org/10.1109/TIP.2006.881959>
- [3] Das, K., Jiang, J., Rao, J.N.K.: Mean squared error of empirical predictor. *The Annals of Statistics* **32**(2), 818–840 (2004) <https://doi.org/10.1214/009053604000000201>
- [4] Huynh-Thu, Q., Ghanbari, M.: Scope of validity of psnr in image/video quality assessment. *Electronics letters* **44**(13), 800–801 (2008) <https://doi.org/10.1049/el:20080522>
- [5] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004) <https://doi.org/10.1109/TIP.2003.819861>
- [6] Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing* **21**(12), 4695–4708 (2012) <https://doi.org/10.1109/TIP.2012.2214050>
- [7] Mittal, A., Soundararajan, R., Bovik, A.C.: Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters* **20**(3), 209–212 (2013) <https://doi.org/10.1109/LSP.2012.2227726>
- [8] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6629–6640 (2017). <https://doi.org/10.5555/3295222.3295408>
- [9] Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: *2006 IEEE Computer Society Conference on Computer*

Vision and Pattern Recognition (CVPR'06), vol. 2, pp. 1735–1742 (2006). <https://doi.org/10.1109/CVPR.2006.100>

- [10] Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3733–3742 (2018). <https://doi.org/10.1109/CVPR.2018.00393>
- [11] Zhang, L., Zhang, L., Mou, X., Zhang, D.: Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing* **20**(8), 2378–2386 (2011) <https://doi.org/10.1109/TIP.2011.2109730>
- [12] Qu, G., Zhang, D., Yan, P.: Information measure for performance of image fusion. *Electronics letters* **38**(7), 1 (2002)
- [13] Haghigiat, M.B.A., Aghagolzadeh, A., Seyedarabi, H.: A non-reference image fusion metric based on mutual information of image features. *Computers & Electrical Engineering* **37**(5), 744–756 (2011) <https://doi.org/10.1016/j.compeleceng.2011.07.012>. Special Issue on Image Processing
- [14] Piella, G., Heijmans, H.: A new quality metric for image fusion. In: Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), vol. 3, p. 173 (2003). <https://doi.org/10.1109/ICIP.2003.1247209>
- [15] Han, Y., Cai, Y., Cao, Y., Xu, X.: A new image fusion performance metric based on visual information fidelity. *Information Fusion* **14**(2), 127–135 (2013) <https://doi.org/10.1016/j.inffus.2011.08.002>
- [16] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 586–595 (2018). <https://doi.org/10.1109/CVPR.2018.00068>
- [17] Moorthy, A.K., Bovik, A.C.: A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters* **17**(5), 513–516 (2010) <https://doi.org/10.1109/LSP.2010.2043888>
- [18] N, V., D, P., Bh, M.C., Channappayya, S.S., Medasani, S.S.: Blind image quality evaluation using perception based features. In: 2015 Twenty First National Conference on Communications (NCC), pp. 1–6 (2015). <https://doi.org/10.1109/NCC.2015.7084843>
- [19] Madhusudana, P.C., Soundararajan, R.: Subjective and objective quality assessment of stitched images for virtual reality. *IEEE Transactions on Image Processing* **28**(11), 5620–5635 (2019) <https://doi.org/10.1109/TIP.2019.2921858>

- [20] Zhou, X., Zhang, H., Wang, Y.: A multi-image stitching method and quality evaluation. In: 2017 4th International Conference on Information Science and Control Engineering (ICISCE), pp. 46–50 (2017). <https://doi.org/10.1109/ICISCE.2017.8050502>
- [21] Cheung, G., Yang, L., Tan, Z., Huang, Z.: A content-aware metric for stitched panoramic image quality assessment. In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), pp. 2487–249 (2017). <https://doi.org/10.1109/ICCVW.2017.793>
- [22] Ling, S., Cheung, G., Le Callet, P.: No-reference quality assessment for stitched panoramic images using convolutional sparse coding and compound feature selection. In: 2018 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6 (2018). <https://doi.org/10.1109/ICME.2018.8486545>
- [23] Wu, C., Wu, F., Qi, T., Huang, Y.: Noisytune: A little noise can help you finetune pretrained language models better. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 680–685 (2022). <https://doi.org/10.18653/v1/2022.acl-short.76>
- [24] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826 (2016). <https://doi.org/10.1109/CVPR.2016.308>
- [25] Chen, X., He, K.: Exploring simple siamese representation learning. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15745–15753 (2021). <https://doi.org/10.1109/CVPR46437.2021.01549>
- [26] Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.: Deep rectangling for image stitching: A learning baseline. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5730–5738 (2022). <https://doi.org/10.1109/CVPR52688.2022.00565>
- [27] Zaragoza, J., Chin, T.-J., Tran, Q.-H., Brown, M.S., Suter, D.: As-projective-as-possible image stitching with moving dlt. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**(7), 1285–1298 (2014) <https://doi.org/10.1109/TPAMI.2013.247>
- [28] Chang, C.-H., Sato, Y., Chuang, Y.-Y.: Shape-preserving half-projective warps for image stitching. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3254–3261 (2014). <https://doi.org/10.1109/CVPR.2014.422>
- [29] Lin, C.-C., Pankanti, S.U., Ramamurthy, K.N., Aravkin, A.Y.: Adaptive as-natural-as-possible image stitching. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1155–1163 (2015). <https://doi.org/10.1109/CVPR.2015.7298719>

- [30] Liao, T., Li, N.: Single-perspective warps in natural image stitching. *IEEE Transactions on Image Processing* **29**, 724–735 (2020) <https://doi.org/10.1109/TIP.2019.2934344>
- [31] Nahler, G.: correlation coefficient, pp. 40–41 (2009). <https://doi.org/10.1007/978-3-211-89836-9-304>
- [32] Jan, H., Tomasz, K.: Comparison of values of pearson’s and spearman’s correlation coefficients on the same sets of data. *Quaestiones Geographicae* **30**(2), 87–93 (2011) <https://doi.org/10.2478/v10117-011-0021-1>