# BackFlip: The Impact of Local and Global Data Augmentations on Artistic Image Aesthetic Assessment

Ombretta Strafforello*, Gonzalo Muradas Odriozola*, Fatemeh Behrad*, Li-Wei Chen*, Anne-Sofie Maerten*, Derya Soydaner*, and Johan Wagemans

KU Leuven, Leuven, Belgium
https://gestaltrevision.be

**Abstract.** Assessing the aesthetic quality of artistic images presents unique challenges due to the subjective nature of aesthetics and the complex visual characteristics inherent to artworks. Basic data augmentation techniques commonly applied to natural images in computer vision may not be suitable for art images in aesthetic evaluation tasks, as they can change the composition of the art images. In this paper, we explore the impact of local and global data augmentation techniques on artistic image aesthetic assessment (IAA). We introduce *BackFlip*, a local data augmentation technique designed specifically for artistic IAA. We evaluate the performance of BackFlip across three artistic image datasets and four neural network architectures, comparing it with the commonly used data augmentation techniques. Then, we analyze the effects of components within the BackFlip pipeline through an ablation study. Our findings demonstrate that local augmentations, such as BackFlip, tend to outperform global augmentations on artistic IAA in most cases, probably because they do not perturb the composition of the art images. These results emphasize the importance of considering both local and global augmentations in future computational aesthetics research.

## 1 Introduction

Evaluating image aesthetics is a subjective task for humans, making it even more challenging for neural networks to perform accurately. This task, known as Image Aesthetic Assessment (IAA) in computer science, is part of the interdisciplinary field of computational aesthetics and involves modelling aesthetic scores. The typical approach involves either binary classification, which classifies an image as low or high in aesthetics [6, 13, 25, 32], or regression, which predicts a continuous aesthetic score for a given image [15, 19, 23, 29, 34]. In the literature on automated IAA [9, 39], deep learning plays a crucial role based on its significant impact across various fields. However, datasets collected for this task are often limited and struggle to reflect the true aesthetic nature of images. Numerous

* Equal contribution.

psychological factors influence aesthetic judgments, resulting in diverse individual preferences. These factors range widely, from low-level image properties and mid-level organizational qualities to high-level semantic content, including individual, social, and cultural factors. Consequently, IAA presents a difficult challenge for artificial intelligence, as it seeks to emulate human aesthetic evaluation. This difficulty is amplified when considering the aesthetic assessment of artworks. Artworks are inherently complex and diverse, characterized by variable compositions and styles (from highly realistic to purely abstract). This specific task, known as Artistic IAA, has yet to be fully explored.

Neural network approaches in computer vision commonly use data augmentation techniques to improve performance. However, in artistic IAA, these techniques exhibit limited effectiveness given the importance of overall composition. When images are modified with data augmentation techniques like cropping, flipping, or color adjustment, the visual aspects that contribute to their aesthetic appeal can be altered. These modifications likely disrupt the original composition, harmony, or emotional impact intended by the artist, invalidating the use of the aesthetic scores originally assigned by human participants.

To address this issue, we examine the effects of data augmentation on artistic IAA. We compare well-known techniques and propose a novel technique called *BackFlip*[1], which involves the *local flipping of image regions*. We first segment the images using the Segment Anything (SAM) model [18], then inpaint the *back*ground, and *flip* the selected segment to implement local data augmentation (see Section 3). Our approach aims to minimize alterations to the overall composition, thereby preserving human aesthetic appreciation, while effectively modifying the visual patterns crucial for computer vision recognition. In Fig. 1, we exemplify global and local image transformations applied to art images, highlighting the distinct impacts of different augmentations. Our study examines the impact of local and global data augmentation techniques on artistic IAA using three benchmark datasets composed of paintings. We emphasize the challenges of artwork datasets in the context of data augmentation in computer vision.

## 2 Related Work

### 2.1 Artistic Image Aesthetic Assessment

Early approaches to artistic IAA involve studies that extract features from paintings for classification. For example, Amirshahi *et al.* [1] uses a set of color features in the field of computer vision and image processing, while Li *et al.* [22] employs features representing both global and local characteristics of a painting. Additionally, Guo *et al.* [10] evaluates visual complexity of paintings using features that capture both global and local aspects.

Recent studies include deep learning approaches such as using convolutional neural networks (CNNs) to predict the aesthetics of Chinese ink paintings [40]. Wilber *et al.* [36] presented a large-scale dataset of contemporary artworks and

[1] The code is available at https://github.com/GMuradas99/BackFlip.

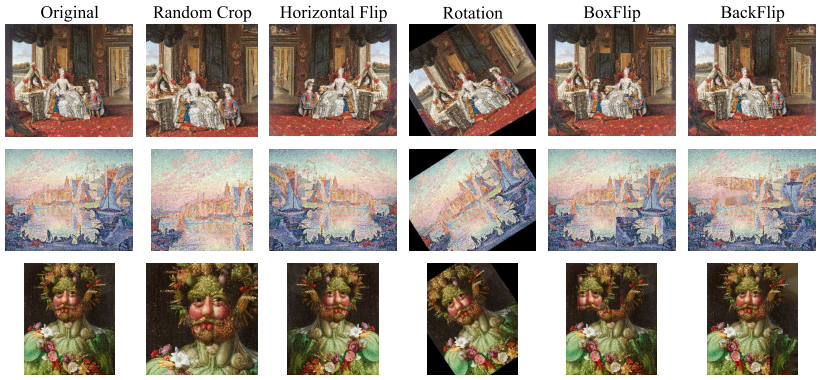


**Fig. 1:** Global (*Random Crop*, *Horizontal Flip*, *Rotation*) and local (*BoxFlip*, *BackFlip*) image transformations on art images from the JenAesthetics dataset [2–4]. The local data augmentations generally preserve the global composition of the images, while introducing considerable pixel-level changes that are often less perceptible to the human eye unless if they distort perceptually important shapes and objects like faces.

used it for artistic style prediction, improving the generality of existing object classifiers, and studying visual domain adaptation. More specifically for the artistic IAA task, the Theme-Style-Color Guided Artistic Image Aesthetics Assessment Network (TSC-Net) [35] assesses art images by fusing aesthetic information with image theme, style, and color. Shi *et al.* [28] presented semantic and style based multiple reference learning for artistic and general IAA. Another recent model, the Style-specific Art Assessment Network (SAAN) [38], evaluates artistic images by combining style-specific and generic aesthetic features. In our study, we adopt deep learning models, including SAAN, to predict aesthetic scores.

### 2.2 Image Data Augmentations

Whether the task involves natural images or artworks, data augmentation techniques are usually necessary for computer vision [21]. Limited labeled data can lead to overfitting. Additionally, labeling data is time-consuming and expensive. To address overfitting, various generalization techniques have been proposed, such as dropout [30] and batch normalization [14]. Among these, data augmentation is the easiest and one of the most common methods to reduce overfitting [20]. Basic data augmentation techniques involve image transformations such as rotation, flipping, and cropping. In the context of artistic IAA, several studies examine data augmentation techniques. For instance, a stacking ensemble method for art style recognition has been presented and the effects of data augmentation such as brightness change and rotation have been examined [26]. In a similar line of research [38], image augmentations to train self-supervised models have been explored. Both methods use global image transformations.

Different from previous work on data augmentation on art images, we explore a *local* data augmentation strategy that does not alter the overall composition of

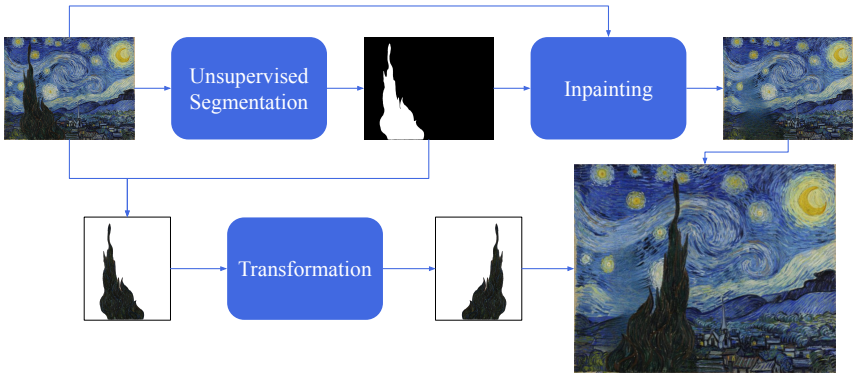


**Fig. 2:** Visualization of the BackFlip pipeline. First, we segment regions in images and inpaint the *back*ground, and then we locally *flip* the selected segment to implement data augmentation.

an artistic image. Local data augmentation has been applied on natural images in various domains but often with highly noticeable visual effects [16,42]. In random erasing [42], for instance, a rectangular part of an image is selected and the pixels are replaced with ImageNet mean values to introduce occlusions. This technique has been shown to complement existing global data augmentation techniques for image classification, object detection, and person re-identification. Another local augmentation technique divides images into rectangular patches and shuffles and augments a selection of patches [16]. This technique has been proposed to exploit the local bias property of CNNs, stating that local augmentations create more diversity relevant to models that extract local features. This approach demonstrates competitive performance with other data augmentation techniques on image classification. Close to this approach, we also suggest *BoxFlip*, where a rectangular patch in an image is augmented. We compare this baseline to our newly proposed BackFlip, which we test under various conditions. We hypothesize that IAA could benefit from local data augmentation more than global augmentation given the importance of composition in artworks for IAA. Our novel technique BackFlip extends previous work by locally transforming image *segments* in order to maintain as much of the overall composition as possible.

## 3 BackFlip

The BackFlip algorithm consists of three primary operations: unsupervised segmentation, inpainting, and local transformations, as shown in Figure 2.

### 3.1 Segmentation

First, we segment the images using the SAM model [18]. We use unsupervised segmentation to accommodate the lack of conventional object classes in abstract art. The SAM implementation in BackFlip returns the segments as binary masks. We exclude segments whose bounding box is larger than 90% of the image area, given that these segments would alter the overall composition when augmented and can no longer be considered local augmentations (e.g., the background). After excluding those segments, the remaining segments are ordered in descending size. One of the hyperparameters of BackFlip is the number of segments $n$ to save. In our tests, SAM detects around 60 segments per artwork on average for all tested datasets.

It should be noted that classical image augmentation techniques are typically applied during training to introduce various random changes to the data at each epoch. However, BackFlip employs the SAM model for segmentation, which would drastically increase the training time when implemented on each epoch (while yielding the same results every epoch). Therefore, the dataset is pre-segmented once before training to optimize computational resources. During training, segments are selected and locally augmented with a given probability on each epoch, ensuring the model sees a wide variety of augmented images.

### 3.2 Inpainting

In the next step, we erase the chosen segments in the images and inpaint the background. BackFlip employs three types of inpainting methods, with various computational costs and different levels of complexity. These methods typically involve a trade-off between image quality and computational efficiency. We present them in descending order of complexity, starting with methods that produce the most realistic inpainted images.

The first method is LaMa [31], a deep learning model that uses fast Fourier convolutions [7], providing a receptive field that covers the entire image while remaining computationally efficient. This approach is significantly more lightweight compared to other state-of-the-art inpainting models based on generative methods like CoModGAN [41] or Stable-Diffusion [27]. The model receives the original image and the segmented area as input and returns an image where the segments are erased and inpainted. Since this method repeats the local statistics of the edge around the erased segment, we first dilate the segment mask to ensure the background no longer contains pixels of the removed segment. Similar to SAM, employing LaMa on each epoch would drastically increase the training time while still yielding the same result on each epoch. Therefore, BackFlip pre-inpaints the images based on their pre-computed segmentation masks before training when LaMa is used. The other inpainting methods demand less computational resources and are therefore implemented during training in BackFlip.

The second group of inpainting methods employs classical computer vision algorithms for efficient real-time data augmentation. We consider two techniques:

Fast Marching Method (Telea) [33], and fluid dynamics (NS) [5]. These methods inpaint image pixels using image gradients or the Laplacian. Treating image intensity as an incompressible flow in fluid dynamics, NS transports the image Laplacian as vorticity into the inpainting area. On the other hand, Telea propagates smoothness along image gradients, iteratively inpainting the image by averaging values from neighboring pixels.

Our final inpainting method is based on the mean or median color of the segment boundary. We first compute a dilated version of the segment mask, from which we subtract the original mask. As such, we obtain the segment boundary (or contour), which is then used to calculate its mean or median color to fill the area of the original segment.

### 3.3 Local Transformations

In the final step, we introduce local transformations in the images. BackFlip offers common data augmentation techniques locally such as vertical and horizontal flipping, random rotation, brightness jitter, downscaling and upscaling. The transformed element is then inserted into the inpainted image. Augmentations are applied on every epoch with a given probability, which is another hyperparameter that can be adjusted.

## 4 Results

In this section, we evaluate the impact of local and global data augmentations on artistic IAA. We assess the performance of BackFlip across three artistic image datasets and four neural network architectures, comparing it with commonly used data augmentation techniques. The models tested are ResNet-18 [11], ResNet-50, ResNeXt-50 [37], and SAAN [38]. We consider a fixed hyperparameter and training setup for each dataset and model combination, with all models pre-trained on ImageNet [8]. We detail the experimental setup in Section 4.1 and present the results in Section 4.2. To ensure robustness and fairness in comparing augmentations, we perform five independent runs for each experiment. Finally, we evaluate the components of the BackFlip pipeline through an ablation study in Section 4.3. Our evaluations are based on the Pearson correlation coefficient (PCC) and Spearman's rank correlation coefficient (SRCC) between the ground-truth aesthetic scores of images and the model's predictions. We also assess the classification performance of the models by defining a threshold of 0.5.

### 4.1 Datasets and Experimental Setup

**BAID.** The Boldbrush Artistic Image Dataset (BAID) [38] consists of 60,337 artistic images covering various art forms, with more than 360,000 votes from online users. Yi *et al.* [38] constructed this dataset entirely from artworks obtained from the website Boldbrush[2]. This website hosts a monthly artwork contest

[2] https://faso.com/boldbrush/popular

where certified artists upload their works and receive public votes from online users. The scores of the images in BAID range from 0 to 10, where 0 represents a lower number of votes and 10 is a higher number of votes. BAID is the most recent IAA dataset and is considered the largest of its kind currently available.

We trained all models for 50 epochs, using various batch sizes depending on the model size (ranging from 50 to 512). All models were trained with a learning rate of 0.001 and the Adam optimizer [17], except SAAN, trained with 0.0001 and AdamW optimizer [24]. For the local data augmentations, we selected three segments and used median inpainting, as this method is less costly for the large BAID dataset. We considered horizontal and vertical flips as local augmentations, each with a probability of 0.5. The number of segments $n$ to save in BackFlip is 5 for all experiments. Pre-segmenting the BAID dataset takes 29 hours, 35 minutes, and 12 seconds for 60337 images on 1 A100 GPU.

**JenAesthetics.** The JenAesthetics Subjective Dataset of Aesthetic Paintings [2–4] consists of 1,628 art images. These images are colored oil paintings, all displayed in museums and scanned at high resolution. The dataset covers 11 art periods/styles, including Renaissance, Baroque, and Impressionism, created by 410 artists. This dataset provides aesthetic quality scores (how aesthetic the image is) and beauty scores (how beautiful the image is). The rating scale is continuous, ranging from 1 to 100. Additionally, it includes scores for liking of color, content, composition, knowledge of the artist, and familiarity with the painting. Each painting was evaluated by 19-21 observers. Due to some broken URLs in the original dataset, we obtained 1,576 out of the 1,628 images. The train, validation, and test sets consist of 1,103, 158, and 315 images, respectively.

We trained all models for 60 epochs. We trained SAAN with a batch size of 32, using AdamW as the optimizer and a learning rate of 0.0001. The other models were trained with a batch size of 128, using Adam as the optimizer and a learning rate of 0.001. For the local data augmentations, we used Telea as inpainting method. When using BackFlip, we applied the same configuration as used in BAID. Pre-segmenting the JenAesthetics dataset takes 1 hour, 49 minutes, and 17 seconds for 1584 images on 1 A4500 Laptop GPU.

**TAD66K.** The Theme and Aesthetics Dataset with 66K images (TAD66K) [12] is specifically designed for IAA, containing 66k images. It covers 47 popular themes, which are the most uploaded on the Flickr website from 2008 to 2021. These themes are grouped into seven superthemes, namely, plants, animals, artifacts, colors, humans, landscapes, and others, which were further divided into 47 subthemes. Images of each theme are annotated independently, and each image contains at least 1200 annotations. The annotation score of each image ranges from 1 to 10, representing the lowest aesthetics to the highest aesthetics. They calculated the average value as the ground-truth of the image.

In our study, we focus on the artistic images within the TAD66K dataset, similar to [28]. This subset contains 1431 labeled artistic images. We maintain the original dataset's split, allocating 289 images for testing and 1,142 images

for training. We randomly split the training set, designating 229 images for validation. We trained all models for 100 epochs with a batch size of 128, a learning rate of 0.0001, and the AdamW optimizer. We used median inpainting for our local data augmentations and implemented BackFlip with the same configuration as used in BAID. Pre-segmenting the TAD66K dataset takes 1 hour, 30 minutes, and 40 seconds for 1431 images on 1 A4500 Laptop GPU.

### 4.2 Artistic IAA Models

To assess the impact of data augmentation on artistic IAA, we consider both global and local augmentation techniques. For the global data augmentation techniques, we include horizontal flip, vertical flip, and rotation. For most augmentations, we first resize every image maintaining the original aspect ratio and reducing the shortest side to 224. Then, we crop part of the image to obtain an input of 224×224. We consider both center cropping and random cropping. Additionally, we consider random cropping without resizing beforehand, referred to as *random resized crop*. We include resize and center crop as a baseline for image preprocessing and report the results before adding augmentation techniques.

For the local data augmentation techniques, we implement BoxFlip by selecting a random patch in the image with a min-ratio of 0.3 and a max-ratio of 0.5, which is then flipped either horizontally or vertically. To assess the impact of the local image augmentations using BackFlip, we also propose the *erase and inpaint* method, which removes segments and inpaints the background without augmenting the segment.

Table 1 presents the results on the BAID dataset, showing an overall tendency for local data augmentations to perform slightly better than global ones. In terms of accuracy, erase+inpaint performs the best for ResNet-18, followed by BackFlip. In terms of correlations, all local augmentations (erase+inpaint, BoxFlip, and BackFlip) outperform the others in PCC, whereas there is no significant difference in SRCC, except for random resized crop, which comes to the forefront. For ResNet-50, BoxFlip outperforms the others. In ResNeXt-50, random resized crop is the best in terms of accuracy, but local data augmentations provide competitive correlations. We observe a similar trend in SAAN. Additionally, we compare the average results of all global data augmentation experiments with those of all local data augmentation experiments, which is shown in the right 3 columns. This comparison shows that local augmentations perform better than global ones, except in ResNeXt-50, where they perform similarly.

We repeat these experiments on the other datasets in our study, with Table 2 showing the results for the JenAesthetics dataset. We observe a similar tendency as in the previous results, but it is more evident. In terms of SRCC, the average performance of all local data augmentations outperforms that of global data augmentations in ResNet-18 and ResNet-50. When comparing accuracy across all models, local augmentations usually outperform or perform similarly to global ones. In terms of correlations, BackFlip is superior to the other techniques in SAAN. We also emphasize that JenAesthetics is a better-curated dataset of paintings compared to the others.

| | Augmentation | BAID | | | | | |
|---|---|---|---|---|---|---|---|
| | | Acc. (%) | PCC | SRCC | Acc. (%) | PCC | SRCC |
| ResNet-18 | R., C.C. | 74.9 ± 1.48 | 0.41 ± 0.04 | 0.29 ± 0.03 | | | |
| | R., Random Crop | 70.24 ± 3.81 | 0.30 ± 0.03 | 0.22 ± 0.02 | | | |
| | Random Resized Crop | 72.5 ± 4.64 | 0.29 ± 0.06 | 0.28 ± 0.05 | | | |
| | R., C.C., Horizontal flip | 70.31 ± 1.6 | 0.31 ± 0.03 | 0.21 ± 0.02 | 70.8 ± 0.2 | 0.30 ± 0.0 | 0.23 ± 0.01 |
| | R., C.C., Vertical flip | 70.04 ± 4.09 | 0.31 ± 0.02 | 0.22 ± 0.02 | | | |
| | R., Rotation, C.C. | 70.93 ± 1.94 | 0.31 ± 0.03 | 0.23 ± 0.03 | | | |
| | R., C.C., Erase+Inpaint | 73.35 ± 3.09 | 0.36 ± 0.05 | 0.23 ± 0.04 | | | |
| | R., C.C., BoxFlip | 71.01 ± 3.19 | 0.37 ± 0.03 | 0.22 ± 0.06 | 72.47 ± 0.71 | 0.36 ± 0.01 | 0.24 ± 0.01 |
| | R., C.C., BackFlip | 72.93 ± 1.95 | 0.34 ± 0.03 | 0.22 ± 0.02 | | | |
| ResNet-50 | R., C.C. | 70.14 ± 1.88 | 0.29 ± 0.05 | 0.2 ± 0.05 | | | |
| | R., Random Crop | 72.29 ± 1.26 | 0.29 ± 0.04 | 0.22 ± 0.03 | | | |
| | Random Resized Crop | 72.58 ± 5.64 | 0.28 ± 0.08 | 0.25 ± 0.05 | | | |
| | R., C.C., Horizontal flip | 71.15 ± 5.7 | 0.32 ± 0.05 | 0.24 ± 0.05 | 71.97 ± 0.11 | 0.30 ± 0.0 | 0.24 ± 0.0 |
| | R., C.C., Vertical flip | 72.17 ± 0.85 | 0.33 ± 0.05 | 0.25 ± 0.05 | | | |
| | R., Rotation, C.C. | 71.69 ± 4.84 | 0.3 ± 0.01 | 0.24 ± 0.02 | | | |
| | R., C.C., Erase+Inpaint | 72.59 ± 2.72 | 0.36 ± 0.03 | 0.23 ± 0.04 | | | |
| | R., C.C., BoxFlip | 74.02 ± 0.6 | 0.38 ± 0.02 | 0.24 ± 0.02 | 73.37 ± 0.28 | 0.37 ± 0.0 | 0.25 ± 0.0 |
| | R., C.C., BackFlip | 70.48 ± 4.57 | 0.33 ± 0.05 | 0.22 ± 0.04 | | | |
| ResNeXt50 | R., C.C. | 73.3 ± 3.13 | 0.37 ± 0.05 | 0.27 ± 0.02 | | | |
| | R., Random Crop | 74.49 ± 1.03 | 0.32 ± 0.09 | 0.25 ± 0.09 | | | |
| | Random Resized Crop | 75.13 ± 2.33 | 0.3 ± 0.13 | 0.32 ± 0.04 | | | |
| | R., C.C., Horizontal flip | 73.48 ± 0.82 | 0.34 ± 0.05 | 0.25 ± 0.04 | 73.6 ± 0.24 | 0.30 ± 0.01 | 0.25 ± 0.01 |
| | R., C.C., Vertical flip | 72.6 ± 1.65 | 0.3 ± 0.02 | 0.23 ± 0.01 | | | |
| | R., Rotation, C.C. | 72.32 ± 2.26 | 0.24 ± 0.14 | 0.2 ± 0.07 | | | |
| | R., C.C., Erase+Inpaint | 73.86 ± 1.25 | 0.35 ± 0.03 | 0.25 ± 0.01 | | | |
| | R., C.C., BoxFlip | 73.93 ± 1.35 | 0.35 ± 0.03 | 0.23 ± 0.02 | 73.51 ± 0.22 | 0.34 ± 0.01 | 0.24 ± 0.0 |
| | R., C.C., BackFlip | 72.74 ± 2.02 | 0.32 ± 0.03 | 0.23 ± 0.02 | | | |
| SAAN | R., C.C. | 74.77 ± 1.37 | 0.37 ± 0.03 | 0.3 ± 0.04 | | | |
| | R., Random Crop | 72.4 ± 1.16 | 0.37 ± 0.03 | 0.29 ± 0.03 | | | |
| | Random Resized Crop | 76.46 ± 0.37 | 0.41 ± 0.03 | 0.37 ± 0.02 | | | |
| | R., C.C., Horizontal flip | 72.4 ± 1.7 | 0.37 ± 0.04 | 0.31 ± 0.04 | 73.43 ± 0.37 | 0.38 ± 0.0 | 0.32 ± 0.01 |
| | R., C.C., Vertical flip | 73.95 ± 0.8 | 0.35 ± 0.02 | 0.3 ± 0.03 | | | |
| | R., Rotation, C.C. | 71.95 ± 1.59 | 0.39 ± 0.01 | 0.31 ± 0.01 | | | |
| | R., C.C., Erase+Inpaint | 74.09 ± 1.01 | 0.41 ± 0.02 | 0.32 ± 0.03 | | | |
| | R., C.C., BoxFlip | 73.27 ± 1.74 | 0.37 ± 0.04 | 0.29 ± 0.03 | 73.61 ± 0.14 | 0.39 ± 0.01 | 0.31 ± 0.01 |
| | R., C.C., BackFlip | 73.47 ± 1.05 | 0.4 ± 0.01 | 0.33 ± 0.02 | | | |

**Table 1:** Results on BAID for four models trained with different global and local data augmentation techniques. *R.* and *C.C.* stand, respectively, for *Resize* and *Center Crop*. We report the average accuracy across at least 5 independent runs. The ensemble mean and standard deviation for global and local data augmentations are displayed in the right column. For the local data augmentations, we used median inpainting.

Lastly, Table 3 shows the results on the TAD66K dataset, presenting similar results between the average local and data augmentations. In terms of accuracy, PCC, and SRCC, we observe that rotation outperforms the others in ResNet-18 and ResNet-50, with local augmentations following closely. In the ResNeXt-50 and SAAN, we observe similar results between the local augmentations and the others. However, it is important to note that TAD66K-Art subset includes a diverse set of images, such as a picture of two paintings and an observer (see Fig. 3, the second row).

| | Augmentation | JenAesthetics | | | | | |
|---|---|---|---|---|---|---|---|
| | | Acc. (%) | PCC | SRCC | Acc. (%) | PCC | SRCC |
| ResNet-18 | R., C.C. | 71.43 ± 2.31 | 0.25 ± 0.03 | 0.24 ± 0.02 | | | |
| | R., Random Crop | 71.3 ± 1.58 | 0.16 ± 0.02 | 0.14 ± 0.02 | | | |
| | Random Resized Crop | 73.49 ± 2.5 | 0.23 ± 0.03 | 0.2 ± 0.04 | | | |
| | R., C.C., Horizontal flip | 73.78 ± 1.48 | 0.13 ± 0.02 | 0.09 ± 0.02 | 70.9 ± 0.6 | 0.17 ± 0.01 | 0.15 ± 0.01 |
| | R., C.C., Vertical flip | 69.27 ± 2.06 | 0.15 ± 0.0 | 0.16 ± 0.01 | | | |
| | R., Rotation, C.C. | 66.67 ± 1.63 | 0.16 ± 0.04 | 0.15 ± 0.03 | | | |
| | R., C.C., Erase+Inpaint | 71.81 ± 1.7 | 0.2 ± 0.02 | 0.18 ± 0.02 | | | |
| | R., C.C., BoxFlip | 69.53 ± 1.85 | 0.22 ± 0.01 | 0.21 ± 0.01 | 69.00 ± 2.54 | 0.19 ± 0.02 | 0.19 ± 0.02 |
| | R., C.C., BackFlip | 65.65 ± 2.22 | 0.16 ± 0.02 | 0.17 ± 0.01 | | | |
| ResNet-50 | R., C.C. | 75.62 ± 0.57 | 0.17 ± 0.02 | 0.15 ± 0.02 | | | |
| | R., Random Crop | 75.56 ± 0.87 | 0.22 ± 0.02 | 0.19 ± 0.02 | | | |
| | Random Resized Crop | 75.87 ± 0.5 | 0.17 ± 0.03 | 0.15 ± 0.03 | | | |
| | R., C.C., Horizontal flip | 70.48 ± 2.51 | 0.17 ± 0.04 | 0.14 ± 0.04 | 72.39 ± 0.61 | 0.16 ± 0.01 | 0.14 ± 0.01 |
| | R., C.C., Vertical flip | 69.71 ± 2.25 | 0.13 ± 0.02 | 0.13 ± 0.02 | | | |
| | R., Rotation, C.C. | 70.35 ± 2.55 | 0.13 ± 0.04 | 0.09 ± 0.03 | | | |
| | R., C.C., Erase+Inpaint | 75.81 ± 0.73 | 0.19 ± 0.03 | 0.18 ± 0.02 | | | |
| | R., C.C., BoxFlip | 74.48 ± 0.73 | 0.19 ± 0.02 | 0.15 ± 0.02 | 74.96 ± 0.60 | 0.20 ± 0.01 | 0.18 ± 0.03 |
| | R., C.C., BackFlip | 74.60 ± 1.90 | 0.22 ± 0.05 | 0.22 ± 0.05 | | | |
| ResNeXt50 | R., C.C. | 69.33 ± 3.16 | 0.14 ± 0.07 | 0.13 ± 0.03 | | | |
| | R., Random Crop | 71.55 ± 4.77 | 0.19 ± 0.05 | 0.18 ± 0.04 | | | |
| | Random Resized Crop | 75.81 ± 1.41 | 0.26 ± 0.03 | 0.22 ± 0.04 | | | |
| | R., C.C., Horizontal flip | 68.83 ± 1.46 | 0.16 ± 0.07 | 0.15 ± 0.04 | 72.15 ± 0.5 | 0.19 ± 0.01 | 0.17 ± 0.01 |
| | R., C.C., Vertical flip | 72.83 ± 1.7 | 0.2 ± 0.03 | 0.17 ± 0.02 | | | |
| | R., Rotation, C.C. | 71.75 ± 2.88 | 0.13 ± 0.03 | 0.13 ± 0.03 | | | |
| | R., C.C., Erase+Inpaint | 75.62 ± 2.08 | 0.19 ± 0.02 | 0.16 ± 0.02 | | | |
| | R., C.C., BoxFlip | 74.92 ± 2.74 | 0.12 ± 0.03 | 0.11 ± 0.04 | 75.13 ± 0.34 | 0.17 ± 0.04 | 0.15 ± 0.03 |
| | R., C.C., BackFlip | 74.86 ± 0.73 | 0.20 ± 0.04 | 0.17 ± 0.06 | | | |
| SAAN | R., C.C. | 73.52 ± 1.72 | 0.18 ± 0.05 | 0.22 ± 0.05 | | | |
| | R., Random Crop | 75.43 ± 1.04 | 0.29 ± 0.02 | 0.28 ± 0.02 | | | |
| | Random Resized Crop | 75.49 ± 1.41 | 0.25 ± 0.06 | 0.26 ± 0.04 | | | |
| | R., C.C., Horizontal flip | 74.54 ± 1.09 | 0.2 ± 0.03 | 0.22 ± 0.02 | 75.12 ± 0.08 | 0.25 ± 0.01 | 0.25 ± 0.02 |
| | R., C.C., Vertical flip | 75.11 ± 1.09 | 0.26 ± 0.02 | 0.24 ± 0.02 | | | |
| | R., Rotation, C.C. | 75.05 ± 1.73 | 0.27 ± 0.03 | 0.25 ± 0.02 | | | |
| | R., C.C., Erase+Inpaint | 74.6 ± 1.92 | 0.24 ± 0.04 | 0.22 ± 0.03 | | | |
| | R., C.C., BoxFlip | 74.92 ± 2.36 | 0.22 ± 0.08 | 0.23 ± 0.06 | 74.64 ± 0.21 | 0.26 ± 0.04 | 0.25 ± 0.03 |
| | R., C.C., BackFlip | 74.41 ± 0.86 | 0.31 ± 0.02 | 0.29 ± 0.02 | | | |

**Table 2:** Results on JenAesthetics for different models trained with global and local image data augmentation techniques. *R.* and *C.C.* stand, respectively, for *Resize* and *Center Crop*. We report the average accuracy across 5 independent runs. For the local data augmentations, we used Telea inpainting. The table format follows that of Table 1.

### 4.3 BackFlip Ablation Study

We observed that BackFlip, as well as the erase+inpaint and BoxFlip techniques, perform well. Here, we evaluate the design choices of BackFlip through an ablation study. Specifically, we test the impact of the inpainting method, the number of locally augmented segments, and the type of local segment augmentations.

**Inpainting method.** The inpainting component in BackFlip (Fig. 2) can be crucial in the visual quality of the output image while it is unclear whether the

| | Augmentation | TAD66K (artwork) Acc. (%) | PCC | SRCC | Acc. (%) | PCC | SRCC |
|---|---|---|---|---|---|---|---|
| ResNet-18 | R., C.C. | 52.94 ± 1.83 | 0.1 ± 0.07 | 0.1 ± 0.07 | | | |
| | R., Random Crop | 57.58 ± 2.14 | 0.24 ± 0.05 | 0.23 ± 0.04 | | | |
| | Random Resized Crop | 53.77 ± 3.87 | 0.25 ± 0.03 | 0.22 ± 0.04 | | | |
| | R., C.C., Horizontal flip | 53.83 ± 3.49 | 0.17 ± 0.05 | 0.15 ± 0.06 | 54.88 ± 0.42 | 0.22 ± 0.01 | 0.20 ± 0.01 |
| | R., C.C., Vertical flip | 52.6 ± 3.1 | 0.18 ± 0.04 | 0.16 ± 0.04 | | | |
| | R., Rotation, C.C. | 56.61 ± 2.96 | 0.27 ± 0.03 | 0.25 ± 0.04 | | | |
| | R., C.C., Erase+Inpaint | 53.63 ± 6.03 | 0.22 ± 0.04 | 0.18 ± 0.05 | | | |
| | R., C.C., BoxFlip | 53.84 ± 1.98 | 0.24 ± 0.04 | 0.19 ± 0.03 | 54.21 ± 0.28 | 0.23 ± 0.0 | 0.19 ± 0.0 |
| | R., C.C., BackFlip | 55.16 ± 2.7 | 0.22 ± 0.03 | 0.2 ± 0.03 | | | |
| ResNet-50 | R., C.C. | 58.83 ± 2.75 | 0.3 ± 0.03 | 0.27 ± 0.03 | | | |
| | R., Random Crop | 61.39 ± 3.02 | 0.32 ± 0.02 | 0.3 ± 0.01 | | | |
| | Random Resized Crop | 60.76 ± 3.08 | 0.3 ± 0.03 | 0.3 ± 0.02 | | | |
| | R., C.C., Horizontal flip | 59.93 ± 3.57 | 0.33 ± 0.04 | 0.3 ± 0.05 | 61.07 ± 0.28 | 0.33 ± 0.01 | 0.31 ± 0.01 |
| | R., C.C., Vertical flip | 59.93 ± 1.71 | 0.3 ± 0.03 | 0.28 ± 0.02 | | | |
| | R., Rotation, C.C. | 63.32 ± 1.45 | 0.38 ± 0.01 | 0.38 ± 0.02 | | | |
| | R., C.C., Erase+Inpaint | 61.53 ± 1.37 | 0.34 ± 0.02 | 0.3 ± 0.01 | | | |
| | R., C.C., BoxFlip | 61.52 ± 1.98 | 0.33 ± 0.02 | 0.32 ± 0.02 | 60.72 ± 0.47 | 0.34 ± 0.0 | 0.32 ± 0.0 |
| | R., C.C., BackFlip | 59.1 ± 1.65 | 0.35 ± 0.03 | 0.32 ± 0.03 | | | |
| ResNeXt50 | R., C.C. | 59.17 ± 3.69 | 0.37 ± 0.04 | 0.34 ± 0.05 | | | |
| | R., Random Crop | 62.77 ± 4.54 | 0.38 ± 0.02 | 0.37 ± 0.01 | | | |
| | Random Resized Crop | 60.49 ± 2.83 | 0.37 ± 0.03 | 0.35 ± 0.02 | | | |
| | R., C.C., Horizontal flip | 55.25 ± 6.9 | 0.39 ± 0.03 | 0.36 ± 0.02 | 59.62 ± 0.56 | 0.38 ± 0.0 | 0.35 ± 0.0 |
| | R., C.C., Vertical flip | 60.55 ± 2.93 | 0.36 ± 0.03 | 0.34 ± 0.02 | | | |
| | R., Rotation, C.C. | 59.03 ± 2.08 | 0.38 ± 0.04 | 0.35 ± 0.03 | | | |
| | R., C.C., Erase+Inpaint | 59.03 ± 1.5 | 0.38 ± 0.02 | 0.34 ± 0.02 | | | |
| | R., C.C., BoxFlip | 60.07 ± 2.97 | 0.39 ± 0.05 | 0.36 ± 0.05 | 59.61 ± 0.18 | 0.38 ± 0.0 | 0.34 ± 0.0 |
| | R., C.C., BackFlip | 59.72 ± 1.13 | 0.37 ± 0.05 | 0.33 ± 0.06 | | | |
| SAAN | R., C.C. | 59.72 ± 2.68 | 0.26 ± 0.05 | 0.28 ± 0.03 | | | |
| | R., Random Crop | 59.31 ± 2.79 | 0.22 ± 0.06 | 0.26 ± 0.03 | | | |
| | Random Resized Crop | 59.58 ± 2.33 | 0.21 ± 0.06 | 0.27 ± 0.03 | | | |
| | R., C.C., Horizontal flip | 58.69 ± 1.95 | 0.21 ± 0.09 | 0.24 ± 0.06 | 59.29 ± 0.13 | 0.21 ± 0.01 | 0.26 ± 0.0 |
| | R., C.C., Vertical flip | 58.69 ± 1.67 | 0.17 ± 0.07 | 0.24 ± 0.03 | | | |
| | R., Rotation, C.C. | 60.21 ± 2.54 | 0.25 ± 0.06 | 0.28 ± 0.04 | | | |
| | R., C.C., Erase+Inpaint | 58.13 ± 1.2 | 0.26 ± 0.06 | 0.26 ± 0.03 | | | |
| | R., C.C., BoxFlip | 60.41 ± 2.62 | 0.23 ± 0.06 | 0.27 ± 0.05 | 59.03 ± 0.41 | 0.25 ± 0.01 | 0.27 ± 0.0 |
| | R., C.C., BackFlip | 58.55 ± 2.21 | 0.26 ± 0.03 | 0.26 ± 0.03 | | | |

**Table 3:** Results on TAD66K - Art for different models trained with global and local image data augmentation techniques. *R.* and *C.C.* stand, respectively, for *Resize* and *Center Crop*. We report the average accuracy across 5 independent runs. For the local data augmentations, we used median inpainting. The table format follows that of Table 1.

models benefit from improved inpainting methods. To showcase the inpainting technique in isolation, we tested BackFlip without inserting a segmented image, which is equivalent to 'Erase + Inpaint' method, in Figure 3. We compared mean, median, NS, Telea, and LaMa inpainting techniques on two example images from TAD66K - Art. The original images in the first column of this figure also exemplify the diversity of images in the 'art' category in TAD66K. One example is a museum picture showing two paintings from an angle, with a visitor partially

occluding one painting. The image as a whole is not a painting; segmenting out the person does more than changing the 'artwork'.

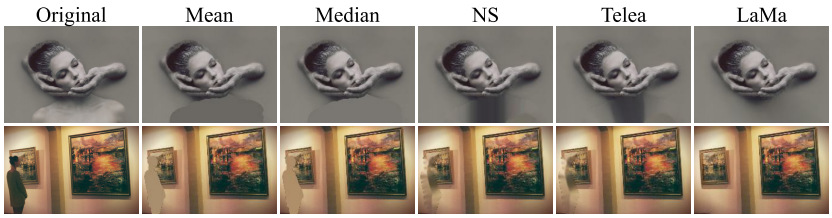


**Fig. 3:** Erase + Inpaint (BackFlip without inserting a segmented image) with different inpainting methods on images from TAD66K - Art.

To compare the inpainting techniques, we train SAAN [38] for artistic IAA on the TAD66K dataset. Table 4 shows the results, keeping the BackFlip setup constant. In this case, we augmented three local segments and used horizontal flipping as the local augmentation. The more refined inpainting techniques, such as LaMa and Telea, yield slightly better results than the others. However, given their higher computational cost and relatively small performance gain, we argue that a simpler approach, such as median inpainting, is a more optimal choice, especially when training on larger datasets.

| **Inpainting method** | **Acc. (%)** | **PCC** | **SRCC** |
|---|---|---|---|
| Mean | $56.33 \pm 1.05$ | $0.28 \pm 0.05$ | $0.26 \pm 0.04$ |
| Median | $57.02 \pm 0.79$ | $0.32 \pm 0.02$ | $0.28 \pm 0.02$ |
| Telea | $\mathbf{59.03 \pm 1.46}$ | $0.33 \pm 0.02$ | $0.31 \pm 0.03$ |
| NS | $56.81 \pm 2.22$ | $0.32 \pm 0.03$ | $0.28 \pm 0.03$ |
| LaMa | $58.48 \pm 1.49$ | $\mathbf{0.34 \pm 0.04}$ | $\mathbf{0.32 \pm 0.03}$ |

**Table 4:** Testing the effect of BackFlip with different inpainting methods on TAD66K - Art using SAAN [38]. We report the average across 5 independent runs.

**Local augmentation types.** Understanding how different image transformations affect perceived aesthetic value for human observers and models is important. We consider six local image transformations using BackFlip, illustrated on example images from TAD66k - Art in Figure 4. We compare the results to assess the effect of BackFlip with different local augmentation techniques using ResNet-18 in Table 5. According to the results, all the local augmentations perform similarly. However, upscale appears to yield higher PCC and SRCC scores, likely because the augmented segments cover more of the inpainted background, thereby mitigating the loss of information due to inpainting. In terms of accuracy, downscale performs slightly better.

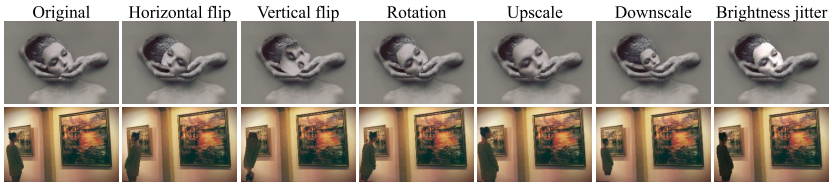


**Fig. 4:** BackFlip with different local transformations on images from TAD66K - Art.

| Local augmentation | Acc. (%) | PCC | SRCC |
|---|---|---|---|
| Horizontal flipping | $53.11 \pm 5.83$ | $0.18 \pm 0.13$ | $0.15 \pm 0.11$ |
| Vertical flipping | $54.33 \pm 5.77$ | $0.19 \pm 0.15$ | $0.17 \pm 0.13$ |
| Hor./Ver. flipping | $54.86 \pm 1.94$ | $0.24 \pm 0.05$ | $0.21 \pm 0.04$ |
| Rotation | $53.34 \pm 5.06$ | $0.20 \pm 0.14$ | $0.18 \pm 0.12$ |
| Upscale | $54.33 \pm 2.45$ | $\mathbf{0.25 \pm 0.02}$ | $\mathbf{0.23 \pm 0.03}$ |
| Downscale | $\mathbf{55.64 \pm 1.28}$ | $0.24 \pm 0.01$ | $0.20 \pm 0.01$ |
| Brightness jitter | $54.19 \pm 3.78$ | $0.22 \pm 0.05$ | $0.19 \pm 0.06$ |

**Table 5:** Testing the effect of BackFlip with different local augmentation methods on TAD66K - Art using ResNet-18. We report the average across 5 independent runs.

**Number of augmented segments.** We illustrate the effect of the number of segments on images from the TAD66k - Art dataset in Figure 5. This figure shows the effects of three local image transformations (rotation, horizontal and vertical flip) using BackFlip, considering up to five augmented segments. As observed, the number of segments augmented through BackFlip significantly affects the visual dissimilarity between the augmented images and the original image.

We compare the effect of the number of augmented segments on model performance in Table 6, keeping every other aspect of the training set-up constant between comparisons. We train ResNet-18, pre-trained on ImageNet, on the TAD66K - Art dataset. Parameters for the BackFlip components and the chosen inpainting method, in this case, LaMa, are fixed. We consistently use either vertical flip or horizontal flip for the local augmentations applied to each segment. The results suggest that one segment is the most optimal choice for local augmentation for artworks. However, we do not observe a significant difference between the results across different numbers of segments, except possibly for four segments. Notably, these results do not show consistent improvements across all five runs. This inconsistency could have multiple explanations. It could be due to varying segment sizes, which change the percentage of image alterations between runs, or the interaction between chosen segments and local augmentations (horizontal vs. vertical flip), influenced by the complex nature of artistic images.

## 5 Conclusion and Future Work

We introduce BackFlip and examine the impact of local and global data augmentation on artistic IAA. Local augmentations, such as BackFlip, preserve the overall composition of images while introducing variations that do not affect


**Fig. 5:** Local image transformations (rotation, horizontal and vertical flip) using BackFlip with increasing number of segments. Images from the TAD66k - Art dataset.

| Number of segments | Acc. (%) | PCC | SRCC |
|---|---|---|---|
| 1 | **55.19 ± 2.44** | **0.26 ± 0.04** | **0.23 ± 0.04** |
| 2 | 54.5 ± 3.42 | 0.24 ± 0.04 | 0.21 ± 0.04 |
| 3 | 54.86 ± 1.94 | 0.24 ± 0.05 | 0.21 ± 0.04 |
| 4 | 52.8 ± 3.06 | 0.24 ± 0.04 | 0.21 ± 0.04 |
| 5 | 54.6 ± 2.48 | 0.23 ± 0.04 | 0.21 ± 0.05 |

**Table 6:** The effect of the number of augmented segments on BackFlip. The results are obtained on the TAD66k - Art dataset using ResNet-18. We report the average across 5 independent runs. As explained in Section 3.1, we exclude the segments covering more than 90% of the image and select the remaining segments in descending size.

global aesthetic qualities, making them advantageous for artistic IAA. Our experiments demonstrate that local augmentations outperform global ones in the majority of our tests. A notable contribution of our study is the inclusion of the erase+inpainting technique within the BackFlip pipeline, as well as BoxFlip, which further enhances the effectiveness of local augmentations. This underscores the importance of local augmentations that preserve overall composition and the crucial role that composition plays in the aesthetics of artworks.

Additionally, we emphasize that the dataset quality plays a crucial role in artistic IAA. A well-curated and diverse dataset is essential for reliable results. We observed that the annotations in some datasets, such as BAID, are too noisy to provide a good supervisory signal. Furthermore, the images in TAD66k, which include pictures of artworks, frames, graffiti, and even nail art, do not always align with the typical global aesthetic qualities expected in paintings.

Our findings suggest that BackFlip is a promising technique for artistic IAA and holds potential for broader applications in aesthetics research. An interesting future work would be to explore the impact of varying parameter settings across different local augmentation methods. Deploying BackFlip can facilitate empirical aesthetics research by collecting ratings from human participants for augmented data, further enriching our understanding of aesthetic evaluation.

**Acknowledgments** Funded by the European Union (ERC AdG, GRAPPA, 101053925, awarded to Johan Wagemans). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. We acknowledge the insightful discussions with Lisa Koßmann and Hayley Hung.

## References


1. Amirshahi, S.A., Denzler, J.: Judging aesthetic quality in paintings based on artistic inspired color features. In: 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA). pp. 1–8. IEEE (2017)
2. Amirshahi, S.A., Hayn-Leichsenring, G.U., Denzler, J., Redies, C.: JenAesthetics subjective dataset: Analyzing paintings by subjective scores. In: Workshop at the European Conference on Computer Vision (2014)
3. Amirshahi, S.A., Redies, C., Denzler, J.: How self-similar are artworks at different levels of spatial resolution? In: Proceedings of the Symposium on Computational Aesthetics. ACM (2013)
4. Amirshahi, S.A., Redies, J.D.C.: JenAesthetics—a public dataset of paintings for aesthetic research (2013), tech. rep., Computer Vision Group, University of Jena Germany
5. Bertalmio, M., Bertozzi, A.L., Sapiro, G.: Navier-stokes, fluid dynamics, and image and video inpainting. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. vol. 1, pp. I–I. IEEE (2001)
6. Chen, H., Shao, F., Mu, B., Jiang, Q.: Image aesthetics assessment with emotion-aware multi-branch network. IEEE Transactions on Instrumentation and Measurement (2024)
7. Chi, L., Jiang, B., Mu, Y.: Fast fourier convolution. Advances in Neural Information Processing Systems **33**, 4479–4488 (2020)
8. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. IEEE (2009)
9. Deng, Y., Loy, C.C., Tang, X.: Image aesthetic assessment: An experimental survey. IEEE Signal Processing Magazine **34**(4), 80–106 (2017)
10. Guo, X., Kurita, T., Asano, C.M., Asano, A.: Visual complexity assessment of painting images. In: 2013 IEEE International Conference on Image Processing. pp. 388–392. IEEE (2013)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
12. He, S., Zhang, Y., Xie, R., Jiang, D., Ming, A.: Rethinking image aesthetics assessment: Models, datasets and benchmarks. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI) (2022)
13. Huang, Y., Li, L., Chen, P., Wu, J., Yang, Y., Li, Y., Shi, G.: Coarse-to-fine image aesthetics assessment with dynamic attribute selection. IEEE Transactions on Multimedia (2024)
14. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pp. 448–456. PMLR (2015)
15. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: MUSIQ: Multi-scale image quality transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5148–5157 (2021)
16. Kim, Y., Uddin, A.S., Bae, S.H.: Local augment: Utilizing local bias property of convolutional neural networks for data augmentation. IEEE Access **9**, 15191–15199 (2021)

17. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
18. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollar, P., Girshick, R.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)
19. Kong, S., Shen, X., Lin, Z., Mech, R., Fowlkes, C.: Photo aesthetics ranking network with attributes and content adaptation. In: Proceedings of the European Conference on Computer Vision. pp. 662–679 (2016)
20. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Communications of the ACM **60**(6), 84–90 (2017)
21. Kumar, T., Mileo, A., Brennan, R., Bendechache, M.: Image data augmentation approaches: A comprehensive survey and future directions. arXiv preprint arXiv:2301.02830 (2023)
22. Li, C., Chen, T.: Aesthetic visual quality assessment of paintings. IEEE Journal of selected topics in Signal Processing **3**(2), 236–252 (2009)
23. Li, L., Huang, Y., Wu, J., Yang, Y., Li, Y., Guo, Y., Shi, G.: Theme-aware visual attribute reasoning for image aesthetics assessment. IEEE Transactions on Circuits and Systems for Video Technology **33**(9), 4798–4811 (2023)
24. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. International conference on learning representations (ICLR) (2019)
25. Lu, X., Lin, Z., Jin, H., Yang, J., Wang, J.Z.: Rapid: Rating pictorial aesthetics using deep learning. In: Proceedings of the 22nd ACM international conference on Multimedia. pp. 457–466 (2014)
26. Menis-Mastromichalakis, O., Sofou, N., Stamou, G.: Deep ensemble art style recognition. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2020)
27. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10684–10695 (June 2022)
28. Shi, T., Chen, C., Li, X., Hao, A.: Semantic and style based multiple reference learning for artistic and general image aesthetic assessment. Neurocomputing **582**, 127434 (2024)
29. Soydaner, D., Wagemans, J.: Multi-task convolutional neural network for image aesthetic assessment. IEEE Access **12**, 4716–4729 (2024)
30. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)
31. Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V.: Resolution-robust large mask inpainting with fourier convolutions. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 2149–2159 (2022)
32. Talebi, H., Milanfar, P.: Nima: Neural image assessment. IEEE transactions on image processing **27**(8), 3998–4011 (2018)
33. Telea, A.: An image inpainting technique based on the fast marching method. Journal of graphics tools **9**(1), 23–34 (2004)
34. Wang, J., Chan, K.C., Loy, C.C.: Exploring clip for assessing the look and feel of images. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2555–2563 (2023)

35. Wang, Y., Cao, W., Sheng, N., Shi, H., Guo, C., Ke, Y.: TSC-Net: Theme-style-color guided artistic image aesthetics assessment network. In: Computer Graphics International Conference. pp. 193–203 (2023)
36. Wilber, M.J., Fang, C., Jin, H., Hertzmann, A., Collomosse, J., Belongie, S.: BAM! the behance artistic media dataset for recognition beyond photography. In: Proceedings of the IEEE international conference on computer vision. pp. 1202–1211 (2017)
37. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1492–1500 (2017)
38. Yi, R., Tian, H., Gu, Z., Lai, Y.K., Rosin, P.L.: Towards artistic image aesthetics assessment: a large-scale dataset and a new method. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22388–22397 (2023)
39. Zhai, G., Min, X.: Perceptual image quality assessment: a survey. Science China Information Sciences **63**, 1–52 (2020)
40. Zhang, J., Miao, Y., Zhang, J., Yu, J.: Inkthetics: A comprehensive computational model for aesthetic evaluation of chinese ink paintings. In: IEEE Access. pp. 225587–225871 (2020)
41. Zhao, S., Cui, J., Sheng, Y., Dong, Y., Liang, X., Chang, E.I., Xu, Y.: Large scale image completion via co-modulated generative adversarial networks. arXiv preprint arXiv:2103.10428 (2021)
42. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 13001–13008 (2020)