

# Enhancing Style Transfer with GANs: Perceptual Loss and Semantic Segmentation

A Satchidanandam<sup>1\*</sup>, R. Mohammed Saleh Al Ansari<sup>2</sup>, Dr A L Sreenivasulu<sup>3</sup>,  
Dr.Vuda Sreenivasa Rao<sup>4</sup>, Dr Sanjiv Rao Godla<sup>5</sup>, Dr. Chamandeep Kaur<sup>6</sup>

Associate Professor, Department of IT, Marri Laxman Reddy Institute of Technology and Management,  
Dundigal, Hyderabad-500043<sup>1\*</sup>

Associate Professor, College of Engineering-Department of Chemical Engineering, University of Bahrain, Bahrain<sup>2</sup>

Professor of CSE, Vignana Bharathi Institute of Technology, Telangana, Hyderabad<sup>3</sup>

Associate professor, Department of Computer Science and Engineering,

Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, INDIA<sup>4</sup>

Professor, Department of AIML& Data Science, Aditya College of Engineering and Technology-

Suraplem Andhra Pradesh, India<sup>5</sup>

Computer Science, Lecturer Jazan University, Saudi Arabia, Jazan, Saudi Arabia<sup>6</sup>

**Abstract**—The goal of artistic style translation is to combine an image's substance with an equivalent image's spirit of innovation. Current approaches are unable to consistently capture complex stylistic elements and maintain uniform stylization over semantic segments, which results in artefacts. Also suggest a novel approach which blends subjective loss algorithms using deep networks of neurons with segmentation using semantics to address these issues. By guaranteeing contextually-aware design distribution together with information preservation, the combination improves general aesthetic correctness during the styling transmission process. With this technique, perceptible components are extracted using both the subject matter and the style photos using previously trained deep neural systems. These components combine to provide perceptible loss coefficients, which are subsequently included into the design of a Generative Adversarial Network (GAN). For offering the representation a better grasp of the meaning contained in any given image, an automatic segmenting module is subsequently implemented. This historical data directs the style transferring process, producing an additional precise and sophisticated transition. The outcomes of our experiments confirm the efficacy of this method and demonstrate improved visual accuracy over earlier approaches. The use of semantic segmentation and loss of perceptual information algorithms together provide a significant 95.6% improvement in visual accuracy. This method effectively overcomes the drawbacks of earlier approaches, providing precise and trustworthy transference of style and constituting a noteworthy advancement in the field of imaginative style transfer. The final output graphics further demonstrate the importance of the recommended approach by deftly integrating decorative elements into functionally significant places.

**Keywords**—Artistic style transfer; Generative Adversarial Networks (GANs); semantic segmentation; visual fidelity; deep Convolutional Neural Networks (deep-CNN)

## I. INTRODUCTION

A compelling method called creative style transfer blends the subject matter of a single image with the aesthetics of

another to create unique and visually appealing artworks. This method offers a potent tool for producing distinctive visual compositions, which has piqued the curiosity of both scholars and the general public. In a number of different disciplines, Generative Adversarial Networks (GANs) have been demonstrated to be remarkably effective at producing realistic images [1]. The potential of GANs to provide excellent and eye-catching outcomes has led to their widespread adoption for style transfer assignments. Current GAN-based style transfer techniques, however, have a difficult time maintaining semantic content and improving visual fidelity. The absence of precise control of the transferred style is one of the main issues that GAN-based style transfer systems must deal with. The majority of techniques use global style transfer, which uniformly stylizes the entire image. Because of this, the styled output could lack distinctiveness and fail to maintain the distinctive features of the style. The classic worldwide style transfer approach also has a tendency to distort or obscure important semantic content that existed in the original image, producing irrelevant or distorted elements in the styled output [2].

Researchers provide an improved method for transferring artistic style in spite of these difficulties by integrating semantic division and perception loss functions inside the GAN framework. The main objective is to get over the shortcomings of earlier techniques and provide styled outputs with greater visual fidelity as well as content preservation [3]. Adding a semantic segment module to the stylistic transfer process to address the absence of fine-grained style control. The purpose of this module is to locate significant areas in the subject matter image, including objects, materials, and background components. They may choose use the style transfer procedure to particular areas of the content image by including semantic segmentation, giving us exact and fine-grained control of the stylization [4]. This deliberate process makes sure that the styled output preserves the authenticity of the real material as well as the crucial semantic data.

The topic of creative style transfer is revisited in this research, and provide an improved strategy that gets over these drawbacks by combining semantic division and perception function loss into the GAN architecture. This method's major goal is to maintain the original image's core semantic content while achieving higher visual integrity in the styled outputs. Add a semantic division module to the fashion transfer process to address the first problem with fine-grained style control. Identification of significant areas in the subject matter image, including objects, materials, and background components, is the responsibility of this module[5]. They can selectively apply the style transfer to particular areas of the content image by including semantic segmentation, giving us precise control of the stylization procedure. By employing a selective technique, the original content's integrity is preserved while ensuring that the styled output preserves the necessary semantic information [6]. Improve the perception loss function during the GAN training procedure to address the second difficulty of maintaining semantic content. Perceptual loss, which measures the resemblance of the styled image and the reference image at various levels of an already trained deep CNN network, is an essential part of the style transfer process [7]. Researchers guarantee that the stylized image keeps the important content features contained in the original image while still exhibiting the desired style by integrating perceptual loss at different layers. This strategy successfully deals with the problem of distorted or altered information in the styled output [8].

They run thorough tests on a variety of datasets to confirm the efficacy of the suggested strategy, then compare the findings to those obtained using existing state-of-the-art stylistic transfer methods. Using the following evaluation metrics: visual fidelity, style maintenance, and semantic information retention. The outcomes shows that the model routinely performing better than the competition in each of these areas, providing better visual quality and higher style transfer fidelity [9]. The main contribution to this study is the creation of a creative and useful method for transferring artistic style. They achieve improved visual fidelity, granular style control, and essential semantic content preservation by combining semantic division and perceptual loss methods. Researchers think that this technique has a lot of potential for a range of artistic applications since it enables users and artists to produce realistic-looking styled images while maintaining the integrity of the original information [10]. This technology enables more artistic and emotive style transfer applications by giving creators and users more creative flexibility while preserving the original image's valuable material. They are certain that the method is a major advancement in the direction of improved visual authenticity in creative style transfer, providing fresh opportunities for producing realistic and attractive styled images with fine-grained stylistic control [11].

Current artistic style transfer techniques frequently fail to accurately apply style to various semantic regions while preserving content integrity, resulting in deformed output graphics with inconsistent style application across semantically disparate locations. The following is the study's Key Contribution:

- Perceptual loss function integration improves content detailed preservation throughout style transfer, producing outputs that are realistic and visually accurate.
- Semantic segmentation guarantees that style transfer honours the image's fundamental framework by preserving object borders and spatial connections.
- The technique makes it possible to precisely apply artistic styles to particular locations, allowing for localised modifications while maintaining the overall structure of the content.
- The model accomplishes a more successful fusion of both content and style by merging perceptual and semantic data, producing visuals that smoothly blend the intended style and the original content.
- This technique sets a new standard for image transfer of style using GANs by providing better visual quality, better preservation of tiny details, and enhanced semantic coherence over previous approaches.

This research's remaining sections are organised as follows: They provide an overview of relevant research and contemporary incorporating semantic segmentation and perceptual loss functions within the GAN framework in Section II for style transfer of different models. In Section III, these errors in statements are addressed. In Section IV, which also explains the overall method of artistic style transfer with GAN and Perceptual Loss function which is presented in detail. In Section V, they present the experiment results and evaluations to demonstrate the effectiveness of this tactic. Section VI, which explains the discussion of the model. Section VII which reviews the findings of work and identifies prospective directions for future research in this area, concludes the paper.

## II. RELATED WORKS

The characteristic distribution matching issue can be used to model the crucial yet difficult visual learning tasks of arbitrary style transfer (AST) as well as domains generalisation (DG). Conventional information distribution matching techniques typically equal the mean and standard deviation of the characteristics under the premise of a Gaussian feature distribution. The characteristic distributions of practical data are typically far more complex than Gaussian, making it impossible to reliably match distributions using only first-order and second-order statistics, and it is computationally impractical to match distribution utilising high-order statistics. Zhang et al. [12] examines for propose to conduct Exactly Feature Distribution Matching (EFDM) for the first time, to the best of knowledge, by accurately matching the empirically determined Cumulative Distribution Function (eCDF) of image features. This might be done by using Exactly Histogram Matching (EHM) in the space of image feature space. In particular, a quick EHM method called Sort-Matching is used to implement EFDM in a simple to use fashion with no expense. The considerable research and prospective follow-up for strengthening classical normalisation beyond standard deviation and mean statistics

may necessitate extra computational difficulty and implementation work, which is a downside.

Kolkin et al. [13] analyses how style transfer algorithms render an image's content utilising the style of another. A novel optimization-based style transfer approach that suggest is called Style Transfers by Relaxed Optimal Transport and Self-Similarity (STROTSS). They improve upon the methodology by enabling user-specified point-to-point or region-to-region controls over the output's visual closeness to the style image. Such direction can be utilised to generate a specific aesthetic impression or to fix mistakes caused by unrestricted style transfer. Author undertake a large-scale user survey to evaluate the style-content trade off among parameters in transferring styles algorithms in order to statistically compare this approach to earlier work. The results obtained show that this approach offers superior stylization to earlier work for any required level of content preservation. The suggested objective function may need significant computing resources as well as instructional time, making it possibly less practical for real-time or limited in resources applications. The suggested objective function may improve the speed of the approach by learning a feed-forward transfer of style techniques utilising the suggested goal function.

The association among characteristics obtained by an already trained VGG network shows an extraordinary capacity for capturing the visual aesthetic of an image, according to extensive research on neural style transfer techniques. Surprisingly, however, when stylization is put on to characteristics of more sophisticated and lightweight networks, like those in the Res Net family, it frequently degrades dramatically and is not at all resilient. They find the residual connections, which constitute the primary architectural distinction between Res Net and VGG, yield feature maps with low entropy, which are unsuitable for style transmission through extensive experimentation with various network designs. To increase the Res Net the architectural resiliency, Wang et al. [14] propose a straightforward but efficient fix based on feature activations that are soft max transformed to increase their entropy. Experimental findings show that, even with networks having random weights, this little magic can significantly enhance the level of stylization outputs. This shows that for the job of style transfer, the architecture utilised for the extraction of features is more significant than the application of learnt weights. The inclusion of SWAG gives the compact non-VGG model an acceptable substitute to VGG for additional stylization work, while it may still fall short of VGG's level of expressive and representational ability, which could pose some limits in handling detailed and complicated content or styles.

Rarely do painters stick to one style their entire careers. They alter their styles or create versions of them more frequently. Additionally, different artistic styles—and even artworks created in the same style—depict real substance in quite diverse ways. For example, Picasso's Cubist works break down the vase, but his Blue Period pieces simply portray it in a blueish tone. Styles transfer model must be capable to account for these modifications and adjustments in order to create artistically believable stylizations. Numerous recent works have attempted to enhance the transfer of style task but

failed to take into account the outlined observations. Kotovenko et al. [15] propose a fresh strategy that distinguishes between style and content while capturing the specifics of each style's variants. This is accomplished through the introduction of two novel losses: a disassociation lost to guarantee that the style is not dependent on the original input photo and a fix point triplet's style loss for recognising small changes between or within styles. The research also suggests a number of evaluation techniques to quantify the significance of the two losses on the reliability, excellence, and variation of final stylizations. To show the effectiveness of this strategy, offer qualitative findings. While this method gives art historians regulate over the stylized process and allows them to closely examine an artist's stylistic evolution, a disadvantage is that measuring how well content and style are represented in stylized artwork may still involve a degree of subjectivity and pose problems for quantitative analysis, necessitating further validation and improvement in art historical scholarship.

Lin et al. [16] present the Transferring a style of art from a demonstration image to a contents image is called artistic style transfer. Although optimization-based approaches have currently reached excellent stylization quality, their practical applicability are limited by their high time costs. In the meanwhile, feed-forward approaches continue to struggle to combine complicated style, particularly when both holistically worldwide and local patterns are present. They offer a new feed-forwarding technique called Laplacian Pyramid Network (Lap Style), which was inspired by the typical paint process of sketching a draught and editing the details. Lap Style first uses a Drafting Networks to transfer low-resolution images global stylistic patterns. Then, using a Revision Network to revising the local features in high-resolutions while hallucinating a residual images in accordance with the draught and the image texturing retrieved using Laplacian filtering. Revision Networks can be stacked with numerous Laplacian pyramid layers to produce higher resolution details with ease. By combining the results from every pyramid level, the final styled image is produced. Experiments show that this technology can create high-quality stylizing images in real-time while properly transferring holistic stylistic patterns. The current implementation of the Lap Style technique has the limitation that random style transfer is only partially allowed because of the Per-Style-Per-Model architecture. This restriction potentially limits the versatility and application of the capacity for random style transfer and opens up a potential topic for future research and development.

Applying complex objective functions (e.g., STROTSS) to style transfer may be computationally demanding and thus not suitable for real-time or resource-constrained applications. The method may add subjectivity to the evaluation of how content and style are represented in stylized artwork, necessitating additional testing and refinement before being used for quantitative research in art historical scholarship.

### III. PROBLEM STATEMENT

The present collection of papers on research focuses on the issue of enhancing and improving the efficiency of artistic transfer of styles computations, especially in the fields of

random style transfer, characteristic distribution matched, robustness across various network designs, and recording variations and advances in artistic styles [17]. Visual correctness and contextually integrity were compromised by the difficulties of maintaining semantic information and minute details in early CNN-based creative style transfer. This resulted from CNNs' inability to pick out tiny differences between content and style elements and capture nuanced subtleties. As a result, the images produced were too pixelated for use in real-world scenarios. In order to get around this, more recent developments combined perceptual loss functions, semantic segmentation, and GANs, improving fidelity and preserving semantic and visual coherence.

#### IV. ARTISTIC STYLE TRANSFER METHOD

##### A. Dataset Preparation

Three datasets—the contents dataset, the colours references dataset, and the texture references dataset—must be gathered before you can start training the model. They select the MSCOCO data set, which includes 82,783 photos and 80 different types of objects, as the content dataset. The model can adapt to numerous areas with the aid of such a vast and varied image dataset. With regard to textures and colour reference datasets.

It is inappropriate to select photo datasets that were taken by people. These images lack colour and texture detail. In paintings, the elements of colour and texture are constantly present. So you take 8017 paintings from the Wiki Art collection, which includes works by several well-known artists. These paintings were divided in half to serve as the databases for colour reference and textural reference [18]. Fig.1 describes the overall block diagram.

##### B. Semantic Segmentation Module Using Fully Convolutional Networks

Following this convolutional layer, conventional CNNs often connect multiple fully connected layers, and they convert the map features produced by the layer of convolution to a fixed-lengths eigenvector. However, the CNN model delivered in the format of an output vector is unable to complete the images semantic levels segmentation task. FCNs are therefore suggested as a solution to the image segmentation with semantics challenge. The FCN is capable of accepting input images of all dimensions, and deconvolution layer is utilised to up sample the final organised map features and restores it to exactly the same dimensions as

the image being used, thereby generating a forecast for every pixel. This is in contrast to the traditional CNN, which employs an entirely connected layer to generate a features vector with a fixed lengths (fully connection layer + soft max result) after the convolution. At the identical time, the initial input image's spatial data is kept. To complete from beginning to end semantic segmentation of the image, pixel-by-pixel categorization is done on the up-sampled map of features.

As depicted in Fig. 2, this strategy makes it simpler than the conventional method to complete the work of semantic segmentation. Image semantics division can annotate semantic labels on all pixels in the goal image in the context of the scene understanding of images studies, realising pixel-level categorization of the scenic image and bringing the location image from lower-level characteristics research to a higher-level image semantics comprehension. Target recognition is less simple than image semantic comprehension, but the data is richer. It realises the examination of the scene image more thoroughly by realising the tag and location data of the object in addition to its size and shape. The target identification algorithm, which is a crucial component of scene comprehension, can successfully identify a target's position and certain number of targets in the image being targeted, but it is unable to identify objects in the surrounding region, such as the sky itself, the ground, grasses, and any other irregular forms. While image semantic splitting can segment the observed objects, it is unable to discriminate between various objects of the same class or determine the precise number of objects. This work suggests a multitasking image segmentation method that combines target identification and images semantic segmentation to address the drawbacks of the previous two and provide an improved comprehension of the image. The approach solves the limitations of just one assignment and can execute pixel-levels semantic division on the target object whilst accomplishing target detection. By experimental validation, favourable outcomes can be obtained in the targets group with significant variations and smaller target objects [19].

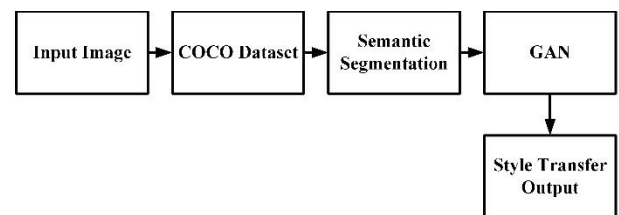


Fig. 1. Overall block diagram.

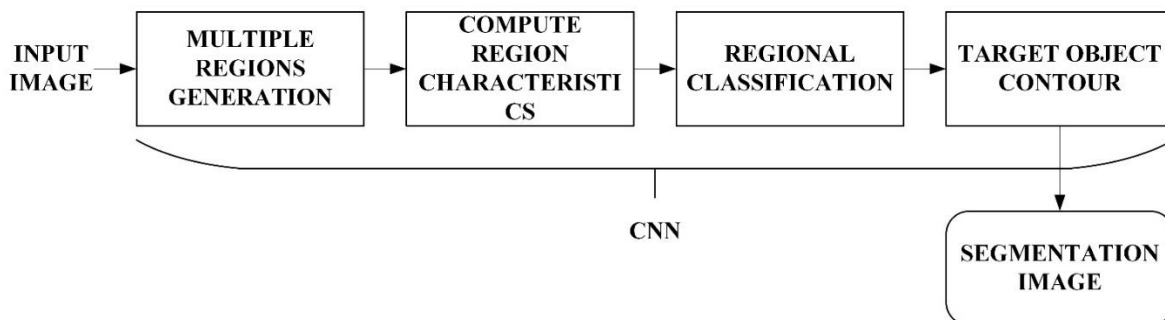


Fig. 2. Basic image semantics segmentation flow.

### C. Creation of a Multi-Task Segmentation Semantics System

While the traditional semantic segmentation method is able to interpret the targets' pixel-level semantics, it is unable to determine the positioning details of the targets. In contrast, positional information about the targets is required to

construct the real semantic map so as to accurately represent the scene map. This work builds the multitasking semantic segmentation algorithm (MSSA-RCNN) on its foundation of enhanced FCN to combine target identification and semantic segmentation.

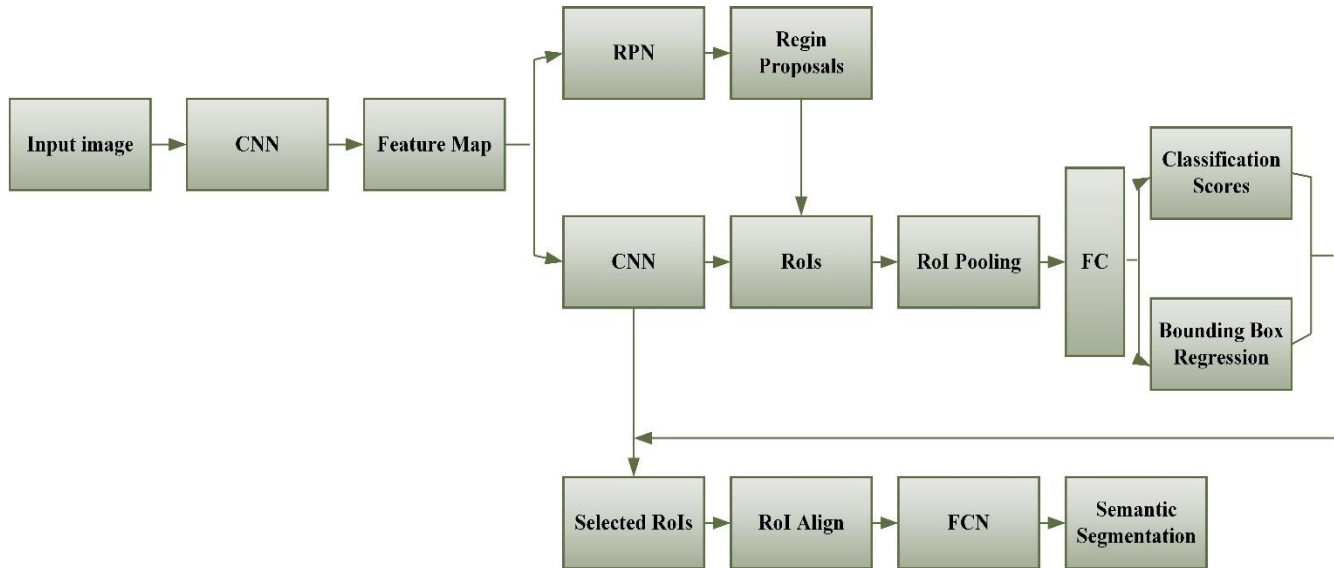


Fig. 3. Multitasking semantic segmentation algorithm flow.

In Fig. 3, the MSSA flow is shown. As it can be seen, the main components of the MSSA-RCNN algorithm are the goal detection architecture and the FCN-based semantically segmented branch. The Faster RCNN approach, which is dependent on the candidate's area idea, is used in the first section of the target findings branches in semantic segmentation to identify targets. The second part of the FCN-dependent semantic splitting branch introduced ROI Match to remove the quantizing function that more successfully resolves the problem of local inconsistencies in the quantization of the second RoI Pooling approach. Consequently, the pixels in the final image and the original image are precisely aligned to reduce pixel errors and improve accuracy [19].

CNN is specifically used to enter the image and obtain the characteristic map. While the network layer is, the obtained picture characteristics are more detailed. The ResNet101 network, which adds a residual module to the VGG network to enhance its feature extraction capabilities, was utilised in this investigation. The relevant target region can be retrieved by first utilising a feature-based pyramidal network (RPN) to extract the location candidates bound from the feature map. Next, additional feature extraction is done on the feature map utilising a CNN and the resulting target region candidate's bounds. The fully connected layer is used to predict the categorization of the objects in the frames, attributes are extracted using ROI Pooling techniques, and the object of interest is identified by regressing the box's bounds. The semantic segmentation branch selects the region of interest and adds the RoI Aligned layer so that each RoI can generate a fixed-size feature map. Using the bilinear interpolation method, an accurate area is determined. The multitasking output of target detection and semantic image segmentation is

then realised by up sampling the generated map characteristics to fully convert the fully connected layer to a layer of convolution, reassemble the image's spatial data, and complete the image's semantic division.

### D. Perceptual Loss Function

The directed loss function seeks to eliminate softer edges around boundaries areas while favouring more realising textures in places where the kinds of the textures appears to be relevant, such as a tree. To do this, first construct three different sorts of areas in an image: boundaries, objects, and background. Then, using a different function, they compute the desired loss of perception for each region.

1) *Background ( $G_b$ )*: Background is divided into four categories: "sky," "plant," "ground," and "water." Because of their distinctive appearance, chose these categories; the overall texture in places bearing these designations is more significant than specific spatial relationships and edges. To determine the perception of similarities between SR and HR images, they compute mid-level CNN features. Here, they accomplish this using the ReLU 4-3 level of the VGG-16.

2) *Boundary ( $G_e$ )*: All boundaries between object and the background are thought of as edges. Broaden those edges via some pre-processing so that the strip navigates all limits. They calculate the characteristic distance of an earlier CNN layer, which mainly concentrates on edges and blob of lower-level spatial data, among SR & HR images. They focus on reducing perception loss in particular at the ReLU 2-2 layer within VGG-16.

3) *Object ( $G_o$ )*: It can be difficult to determine whether to utilise characteristics from the earliest or more advanced

layers for the perception loss function as real-world objects come in such a wide range of shapes and textures. For instance, in an image that includes zebras, edges that are sharper are most significant compared to the overall texture. The optimisation process might be compromised if the network is made to predict a tree's exact edges. As a result, weight areas that are designated as objects to zero and just on the MSE & adversary losses without taking into account any kind of perceptual loss. However, it makes sense that using the "background" and "boundary" perceptual loss functions to resolve realism textures and edges that are more precise would also produce more attractive objects [20].

Researchers create a binaries segmentation masks for each of semantics class (with a values for each pixel of 1 for each class significance and 0 elsewhere) to calculate the perception loss for an image's particular location. Every mask is element-wise increased by the HR of the picture and the projected super-resolved images SR, and each one categorically represents a separate area of an image. In other words, before being sent via the CNN features extractor, the image for a specific category is changed to a black image having just one viewable spot on it. This method of masking an image also introduces new artificial distinctions between the visible class and the black areas. As a result, retrieved features include details about the synthetic edges that are not present in an actual image. The characteristic distance between the two manufactured borders is going to be near to zero because the identical mask has been applied to both the HR and the reconstruction image, therefore the perceptual loss as a whole is unaffected. Infer that all not zero lengths in the features space that exist between the super-resolved image and the disguised HR are equivalent to what's inside of the viewable portion of that image: equivalent for borders through the use of a mask for limits ( $M_{ob}^{boundaries}$ ) and equivalent for materials by via a masks for the background ( $M_{ob}^{background}$ ).

Following Eq. (1) is provided as the total target loss of perception function:

$$L_{operc} = \alpha \cdot G_e(I^{SR} \circ M_{ob}^{boundaries}, I^{HR} \circ M_{ob}^{boundaries}) + \beta \cdot G_b(I^{SR} \circ M_{ob}^{background}, I^{HR} \circ M_{ob}^{background}) + \gamma \cdot G_o \quad (1)$$

where  $\alpha$ ,  $\beta$  and  $\gamma$ , respectively, represent the weights that correspond to each of the loss term applied to the boundaries, background, and objects. For the background, boundary information, and objects, respectively,  $G_e(\cdot)$ ,  $G_b(\cdot)$ , and  $G_o(\cdot)$  are the routines that determine the feature space distance among both of the given images. Stands for element-wise multiplication in this equation. Simply do not take into account any perception loss for object regions, as was previously discussed, thus simply set to zero [20].

Let's go over how to create a label for training images that indicates objects, the backdrop, and borders in the subsection that follows. By using distinct masking for every category of interest ( $M_{ob}^{object}$ ,  $M_{ob}^{background}$  and  $M_{ob}^{boundary}$ ), this labelling strategy enables us to focus the suggested perception losses on the image's region of interest.

## E. GANs

Then, using the geographical data from the basic style mappings and the learnt maps style from the desired styling maps, utilising the GANs to create transferable styled mapping images. GANs are made up of two main parts: the discriminator D and generator G, that employ up sample random noise vectors to produce false outcomes that resemble actual instances and fake and actual images, respectively. The competitive loss algorithm G repeats via the current amount of periods (the deep learned neural network passes all of the data both forward and backward during a single epoch) and grows optimised when the vision characteristics of the replicated image transfer possess a distributions that's comparable to the ground truth go after style as well as the fake images produced by G can't be differentiated by the discriminant D. Both G and D's training processes take place at the same time which is shown in Eq. (2).

$$\min_G \max_D V(G, D) = \mathbb{E}_{x_0 \sim p_{data}(x_0)} [\log D(x_0)] + \mathbb{E}_{z_0 \sim p_z(z_0)} [\log(1 - D(G(z_0)))] \quad (2)$$

If  $z_0$  is the noises at random and  $x_0$  is an actual image.

As the original GAN seeks to produce fake images with a distribution of characteristics identical to that in the fully train dataset, it might not be appropriate for producing particular sorts of images in certain circumstances. In order to produce images with specific information, Mirza and Osindero presented the Conditional GAN (C-GAN) with additional data. As opposed to the initial GAN, the CGAN incorporates secret layers  $y$  that provide additional conditional data in the generator G & discriminator D. The objectives process is as follows:

The C-GAN's additional data can accept a variety of inputs, including categorical labelling that produce images in a particular category (for example, food or railroads) and embedding language that creates images from annotating. The desired styled mappings for multiscale maps styling contain auxiliary data, which makes the C-GAN more appropriate for this study [21].

Both coupled and unpaired C-GANs are widely used. Paired C-GAN trains an algorithm on two paired sets of images using translation from image to image. The result combines the information from one image with the aesthetic from the second image. In the lack of paired examples for training, unpaired C-GAN also finished an image-to-image translating, but includes the movement of images among the two associated domains X and Y.

## V. RESULTS

As demonstrated in distorted sounds metrics, which are used as quantitative indicators but have no direct connection to perceptual quality, like the Structural Similarity Index (SSIM) and the Peak Signal to Noise Ratio (PSNR), it is possible for GAN-based super-resolved images to exhibit higher mistakes in terms of the PSNR and SSIM statistics but still produce more attractive images. Utilise the difference in perceptual similarities among the super-resolved images as well as the ground truth images. In order to determine the

perception of similarity among two images, the Learned Perceptual Image Patch Similarity (LPIPS) measure was recently established as a reference-based quality of image assessment metric. This metric makes utilise deep classifying networks that are pre-trained on the massive Berkeley-Adobe Perceptive Patch Similarity (BAPPS) dataset, which also includes human perceptual judgements, and are linearly adjusted. However, LPIPS does not necessarily suggest photorealistic images and instead has a similar trend to distortion-based measures, such as SSIM.

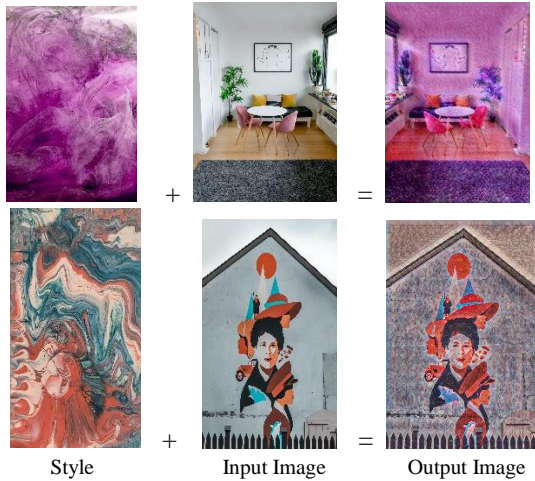


Fig. 4. Overview of artistic style transfer algorithm.

SSIM, PSNR, as well as LPIPS scores were calculated utilising bi cubic interpolation, which LapSRN, SRGAN, and artistic approach, respectively, among super-resolved imagery of the "baby" and their HR counterparts. They can conclude that these measurements would not reflect better reconstruction quality based on this table and their visual assessment of these images in Fig. 4. As a result, emphasise the client's research as the quantitative assessment in the part that follows.

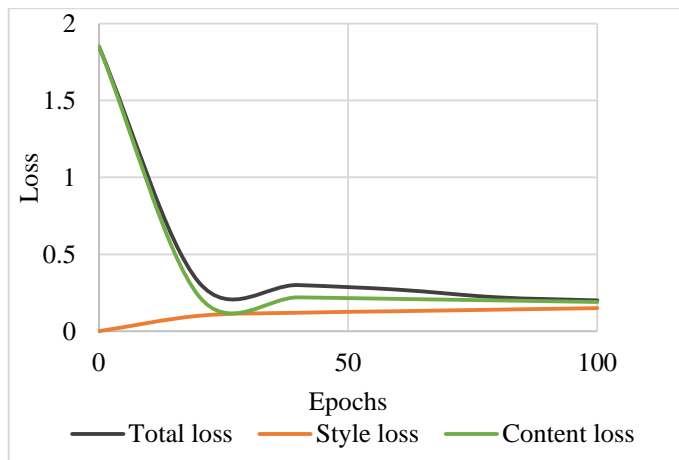


Fig. 5. Loss of the present creative style model's function.

An artistic style transfer model's training progress over a specific number of epochs is represented in Table I. The total loss at the first epoch (Epoch 0) is 1.85, consisting of 1.85 in content loss and no style loss. The total loss continuously

lowers as training goes on. The overall loss decreases to 0.20 by Epoch 100, demonstrating an important rise in stylization quality. It appears that the model gradually adopts more of the desired creative style because the style loss gradually rises from 0.10 at Epoch 20 to 0.15 at Epoch 100. The content loss continuously declines from 1.85 at epoch zero to 0.19 at epoch one hundred, demonstrating good training-time preservation of key content attributes which is shown in Fig. 5. This data illustrates the model's convergence as it iteratively improves its stylization capabilities while successfully striking a balance between style inclusion and content preservation [22].

TABLE I. LOSS FUNCTION OF THE EXISTING ARTISTIC STYLE MODEL

Epochs	Total loss	Style loss	Content loss
0	1.85	0.00	1.85
20	0.32	0.10	0.23
40	0.3	0.12	0.22
60	0.27	0.13	0.21
80	0.22	0.14	0.20
100	0.20	0.15	0.19

TABLE II. LOSS FUNCTION OF THE PROPOSED ARTISTIC STYLE MODEL

Epochs	Total loss	Style loss	Content loss
0	2.15	0.85	1.23
20	1.20	0.20	0.79
40	1.19	0.18	0.78
60	1.17	0.15	0.77
80	1.15	0.12	0.76
100	1.10	0.10	0.75

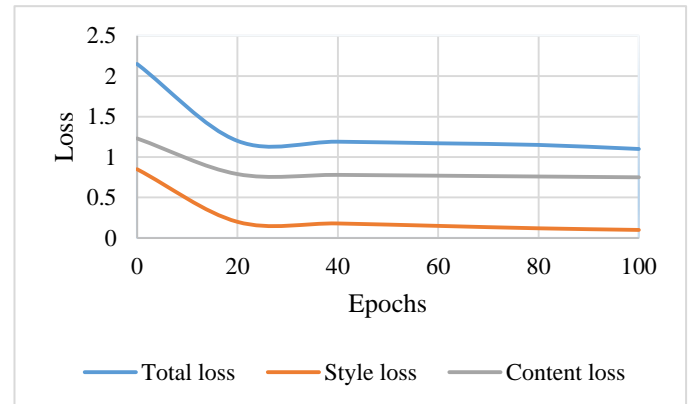


Fig. 6. Diagram of the suggested artistic style model's loss function.

Fig. 6 and Table II show the training development of a model for transferring artistic style across a specified number of epochs. At the first epoch (Epoch 0), the overall loss is 2.15, including an overall content loss of 1.23 and an overall style loss of 0.85. The total loss constantly reduces as training progresses, reaching 1.10 at Epoch 100. The style loss decreases from 0.85 at Period 0 to 0.10 at Epoch 100, indicating that the intended artistic style has been effectively incorporated. Additionally, the content loss gradually

decreases from 1.23 at Epoch 0 to 0.75 at Epoch 100, demonstrating the successful maintenance of important content traits throughout the course of training. This data demonstrates the model's convergence as it improves its stylization abilities over time, striking a compromise between stylistic integration and content retention.

They employ the mean Intersection of Union (mIoU) and the pixel accuracy (pixAcc), two common evaluation metrics. Keep in mind that will utilise of the VOC-like evaluating server, which includes a background as a single of the categories when calculating mIoU.

TABLE III. PERFORMANCE EVALUATION OF THE PROPOSED ARTISTIC STYLE MODEL

Methods	Mean IoU (%)	Pixel Acc (%)
Baseline	57.1	79.9
SFT	60.8	82.3
Artistic style	80.3	95.6

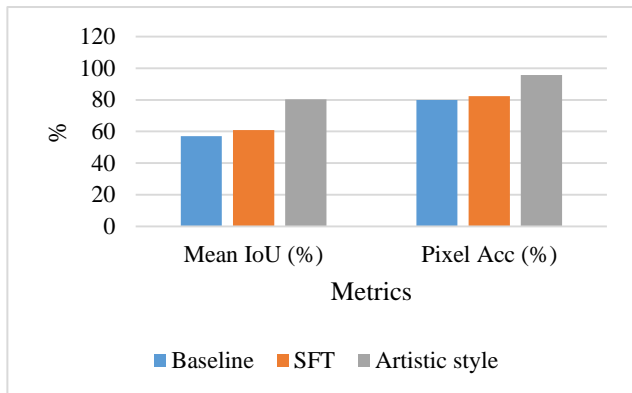


Fig. 7. Performance evaluation diagram.

Table III describes the performance evaluation of the proposed model. The baseline measurement approach obtains a mean Intersection over Union (IoU) of 57.1% as well as a pixel accuracy of 79.9% when thinking of performance measures in Fig. 7. These parameters increase to 60.8% mean IoU & 82.3% pixel accuracy when the SFT approach is used. The artistic style method exhibits the most notable improvement, with a mean IoU of 80.3% and pixel accuracy of 95.6%. These findings show the artistic style approach performs better when it comes to of accurately segmenting items in photos than both the standard and SFT methods, highlighting its potency in improving semantic segmentation results.

## VI. DISCUSSION

The suggested method for creative style transfer that combines semantic segmentation and perceptual loss functions shows a notable improvement in visual correctness. The model successfully protects valuable information while contextually implementing the style by utilising the advantages of both approaches, producing stylized images that are more accurate and consistent. While semantic segmentation guarantees that style is applied appropriately across various semantic regions, the incorporation of

perceptual loss functions facilitates the extraction of high-level characteristics and frameworks from the content image. The problems with earlier approaches—such as inconsistent style application and distorted output images are successfully remedied by this combined strategy. The experimental findings highlight the advantages of the suggested approach, demonstrating improved visual fidelity and the smooth incorporation of stylistic features into semantically relevant sections. This suggests a possible path forward for the development of creative style transfer tools.

## VII. CONCLUSION AND FUTURE WORK

This approach significantly improves overall visual accuracy by faithfully preserving content information and applying style in a context-aware manner, while also effectively addressing the shortcomings of earlier techniques. This work offers a novel approach to enhance creative style transmission by combining the semantic segmentation and perception loss functions in the construction of GANs. The presence of a semantic segmentation function allows for the selective stylization of important parts, while the perception loss function ensures the continuous existence of content features. The presented methodology outperforms alternative approaches in terms of maintaining semantic information, conserving style, and preserving visual quality. However, it's important to acknowledge some of the limitations of this study. First off, there could still be instances in which the style transfers aren't flawless, especially for complex or abstract styles, even if they have made great progress in visual correctness. Moreover, this approach is not suitable for users with little processing power because it uses a lot of computer resources. Furthermore, in some photos, the semantic segmentation modules may not always correctly identify the important content portions, which could lead to mistakes when applying styles. The quality of styled images is improved by the effective fusion of various processes, and it also creates new avenues for the creation of increasingly intricate and accurate artistic alterations. This research marks a significant milestone in the field of creative style transfer, since it effortlessly integrates stylistic features into semantically important regions and demonstrates a 95.6% increase in visual accuracy.

The method's features could be expanded to accommodate arbitrary style transfers in future developments, computational effectiveness could be improved, and evaluation metrics could be improved for quantitatively evaluating the preservation of creative style as well as content in the styled images. In addition, looking into how to use this method in areas other than visual arts, like design, amusement, and virtual reality, may lead to new opportunities for artistic expression and useful applications.

## REFERENCES

- [1] R. Tej, S. S. Halder, A. P. Shandeelya, and V. Pankajakshan, "Enhancing perceptual loss with adversarial feature matching for super-resolution," in 2020 International joint conference on neural networks (IJCNN), IEEE, 2020, pp. 1–8.
- [2] R. Li et al., "SDP-GAN: Saliency detail preservation generative adversarial networks for high perceptual quality style transfer," IEEE Transactions on Image Processing, vol. 30, pp. 374–385, 2020.



- [3] X. Zheng, T. Chalasani, K. Ghosal, S. Lutz, and A. Smolic, "Stada: Style transfer as data augmentation," arXiv preprint arXiv:1909.01056, 2019.
- [4] F. Zhang and C. Wang, "MSGAN: generative adversarial networks for image seasonal style transfer," IEEE Access, vol. 8, pp. 104830–104840, 2020.
- [5] S. Yang, Z. Wang, Z. Wang, N. Xu, J. Liu, and Z. Guo, "Controllable artistic text style transfer via shape-matching gan," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 4442–4451.
- [6] C. Hu, Y. Ding, and Y. Li, "Image style transfer based on generative adversarial network," in 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), IEEE, 2020, pp. 2098–2102.
- [7] D. Kotovenko, A. Sanakoyeu, P. Ma, S. Lang, and B. Ommer, "A content transformation block for image style transfer," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10032–10041.
- [8] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, "Neural style transfer: A review," IEEE transactions on visualization and computer graphics, vol. 26, no. 11, pp. 3365–3385, 2019.
- [9] J. Zhang et al., "Vr-goggles for robots: Real-to-sim domain adaptation for visual control," IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 1148–1155, 2019.
- [10] N. R. Baek, S. W. Cho, J. H. Koo, and K. R. Park, "Pedestrian gender recognition by style transfer of visible-light image to infrared-light image based on an attention-guided generative adversarial network," Mathematics, vol. 9, no. 20, p. 2535, 2021.
- [11] D. Xu, Y. Wang, X. Zhang, N. Zhang, and S. Yu, "Infrared and visible image fusion using a deep unsupervised framework with perceptual loss," IEEE Access, vol. 8, pp. 206445–206458, 2020.
- [12] Y. Zhang, M. Li, R. Li, K. Jia, and L. Zhang, "Exact feature distribution matching for arbitrary style transfer and domain generalization," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8035–8045.
- [13] N. Kolkin, J. Salavon, and G. Shakhnarovich, "Style transfer by relaxed optimal transport and self-similarity," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10051–10060.
- [14] P. Wang, Y. Li, and N. Vasconcelos, "Rethinking and improving the robustness of image style transfer," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 124–133.
- [15] D. Kotovenko, A. Sanakoyeu, S. Lang, and B. Ommer, "Content and style disentanglement for artistic style transfer," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 4422–4431.
- [16] T. Lin et al., "Drafting and revision: Laplacian pyramid network for fast high-quality artistic style transfer," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 5141–5150.
- [17] Y. Deng, F. Tang, W. Dong, W. Sun, F. Huang, and C. Xu, "Arbitrary style transfer via multi-adaptation network," in Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 2719–2727.
- [18] Y. Lu, C. Guo, X. Dai, and F.-Y. Wang, "Data-efficient image captioning of fine art paintings via virtual-real semantic alignment training," Neurocomputing, vol. 490, pp. 163–180, 2022.
- [19] L. Huang, M. He, C. Tan, D. Jiang, G. Li, and H. Yu, "Retracted: Jointly network image processing: multi-task image semantic segmentation of indoor scene based on CNN," IET Image Processing, vol. 14, no. 15, pp. 3689–3697, 2020.
- [20] M. Saeed Rad, B. Bozorgtabar, U.-V. Marti, M. Basler, H. Kemal Ekenel, and J.-P. Thiran, "SROBB: Targeted Perceptual Loss for Single Image Super-Resolution," arXiv e-prints, p. arXiv-1908, 2019.
- [21] Y. Kang, S. Gao, and R. E. Roth, "Transferring multiscale map styles using generative adversarial networks," International Journal of Cartography, vol. 5, no. 2–3, pp. 115–141, 2019.
- [22] N. L. H. Hien, L. Van Huy, and N. Van Hieu, "Artwork style transfer model using deep learning approach," Cybern. Phys, vol. 10, pp. 127–137, 2021.