DOI: 10.1111/cgf.14473
EUROGRAPHICS 2022 / R. Chaine and M. H. Kim
(Guest Editors)
Volume 41 (2022), Number 2

Semantic Segmentation in Art Paintings

N. Cohen[1] and Y. Newman[2] and A. Shamir[3]

[1]The Hebrew University of Jerusalem [2]Tel-Aviv University [3]Reichman University

Figure 1: *Examples of results of our method for semantic segmentation of artistic paintings in various styles – each color represents a class. We use unsupervised domain adaptation on the DRAM (Diverse Realism in Art Movements) dataset we collected. DRAM contains figurative paintings from the Realism, Impressionism, Post-Impressionism and Expressionism art movements (the first two rows). The third row shows examples of segmentation of artistic styles which were unseen during training.*

Abstract

Semantic segmentation is a difficult task even when trained in a supervised manner on photographs. In this paper, we tackle the problem of semantic segmentation of artistic paintings, an even more challenging task because of a much larger diversity in colors, textures, and shapes and because there are no ground truth annotations available for segmentation. We propose an unsupervised method for semantic segmentation of paintings using domain adaptation. Our approach creates a training set of pseudo-paintings in specific artistic styles by using style-transfer on the PASCAL VOC 2012 dataset, and then applies domain confusion between PASCAL VOC 2012 and real paintings. These two steps build on a new dataset we gathered called DRAM (Diverse Realism in Art Movements) composed of figurative art paintings from four movements, which are highly diverse in pattern, color, and geometry. To segment new paintings, we present a composite multi-domain adaptation method that trains on each sub-domain separately and composes their solutions during inference time. Our method provides better segmentation results not only on the specific artistic movements of DRAM, but also on other, unseen ones. We compare our approach to alternative methods and show applications of semantic segmentation in art paintings. The code and models for our approach are publicly available at: https://github.com/Nadavc220/SemanticSegmentationInArtPaintings.

CCS Concepts

• ***Imaging and Video*** → *Image Segmentation; Texture Synthesis;* • ***Methods and Applications*** → *Neural Net;*

1. Introduction

Semantic segmentation of photographs, where each pixel is assigned to one of a set of predefined classes is a difficult task even using today's methods based on neural networks. Methods that train with segmented photographic datasets with around 20 classes, such as PASCAL VOC 2012 [EVW*12] achieve high mean-IOU results [CBP*16; CZP*18; LWLW17; ZSQ*17]. Semantic segmentation becomes even harder in the artistic domain. Artistic paintings have a very different appearance compared to natural photographs, even when concentrating only on figurative art (i.e. non-abstract). They also have much larger diversity in terms of both colors and shapes of objects, and backgrounds.

In this paper, we address the problem of semantic segmentation of (figurative) artistic paintings (see Figure 1). Gathering and annotating an artistic painting dataset is a daunting task, as there are numerous styles and genres within the artistic domain. Hence, our work builds an unsupervised solution using domain adaptation that not only provides a segmentation solution to paintings in some predefined artistic styles, but also allows to segment paintings in unseen styles.

Our method uses two steps. The first step creates a pseudo training-set in some predefined artistic styles (we used Realism, Impressionism, Post-Impressionism and Expressionism) by using style-transfer methods on the existing photographic ground-truth data of PASCAL VOC 2012. We call such datasets *pseudo-paintings* and use them to train basic semantic segmentation networks for these styles. In the second step, we further refine these networks by using a domain confusion technique using PASCAL VOC 2012 as the segmented ground truth source domain and real artistic paintings as the target domain. To segment a new painting, not necessarily from the original domain styles, we first map it to a style latent-space and then combine the segmentations produced by our trained networks based on the similarity of the painting to each domain.

Previous domain adaptation solutions for semantic segmentation mostly concentrate on adapting synthetic rendered images to real photographs (e.g. using GTA5 computer game [RVRK16] to CityScapes dataset [COR*16]). Our work requires the opposite direction – adaptation of real photographs to synthetic, artistic data. We show that simple use of existing domain adaptation techniques does not provide much gain. Even using their original data, reversing the adaptation direction of these methods (i.e. adapting CityScapes to GTA5) reduces the success rate considerably (see supplemental material). This means that adaptation, in general, is *not symmetric* in terms of domains, and there is a need for specialized solutions to adapt photographs to the synthetic domain of paintings.

We see two main reasons for the difficulty of adapting the segmentation of photographs to paintings. The first reason involves the *domain gap*: there are large differences in the characteristics of artistic paintings compared to real photographs. The second reason is *domain diversity*: artistic paintings cannot be seen as a single coherent domain for learning as they encompass a plethora of styles and movements. Our proposed method tackles both challenges. To tackle the domain gap we use style transfer to create pseudo-paintings for training in the first step. To tackle the domain diversity, we separate the target artistic domain to sub-domains and build a *multi-domain adaptation* solution by combining their results during inference.

To guide both the style transfer step and the domain confusion step we use a new dataset we gathered called DRAM: Diverse Realism in Art Movements. DRAM was intentionally created with high variability and large domain gaps. It is comprised of figurative paintings from four art movements: Realism, Impressionism, Post-Impressionism, and Expressionism. These art movements have highly diverse pattern and geometric styles. The objects and scenes painted do not always appear in their true colors and patterns, and their geometric structure is often distorted (see Figure 1). We use DRAM as the target data for adaptation. For our source dataset we chose PASCAL VOC 2012 [EVW*12] as it contains a significant number of classes which are more common in classic artwork.

Figure 2 provides an overview of our approach. We first train a style transfer network on the DRAM data, but use it separately for each sub-domain to create pseudo-paintings of each artistic movement, capturing its unique characteristics. Next, we train semantic segmentation networks using these pseudo-paintings with their original segmentation labels. Lastly, we apply adversarial domain confusion to further refine the segmentation network of each sub-domain using DRAM's real paintings. During inference, given an input painting, we first map it to a style feature-space using Gram matrices [GEB15] as style descriptors. In this space, we find its k-nearest neighbors from the mapped paintings of the DRAM dataset. Based on the ratio of neighbors belonging to each sub-domain we use a weighted combination of the sub-domains segmentation solutions to segment the input painting (see Figure 6). Our experiments show that using such multi-domain inference can be used to segment paintings from unseen artistic domains, but surprisingly it also improves the segmentation results of paintings from the original sub-domains.

To summarize our contributions are:

- We present the first semantic segmentation solution for artistic paintings.
- Our method combines multi-domain adaptation for the highly diverse domain of art paintings.
- We present the DRAM dataset: a new artistic domain adaptation benchmark with a fully segmented test set.

We present results of segmenting artistic paintings, experiments of an ablation study, and show applications of our method. The new DRAM artistic benchmark as well as our code will be released for future research. We see their contribution not only for semantic segmentation of art paintings, but also to domain adaptation development in general, by challenging generalization in highly diverse domains, and testing adaptation from the real domain to a synthetic domain.

2. Related Work

Semantic Segmentation. The leading approach for semantic segmentation of images uses CNNs [RPB15; WZH*19; HGDG17; CBP*16; CZP*18; SZJ*19]. Given an input image, a neural net-

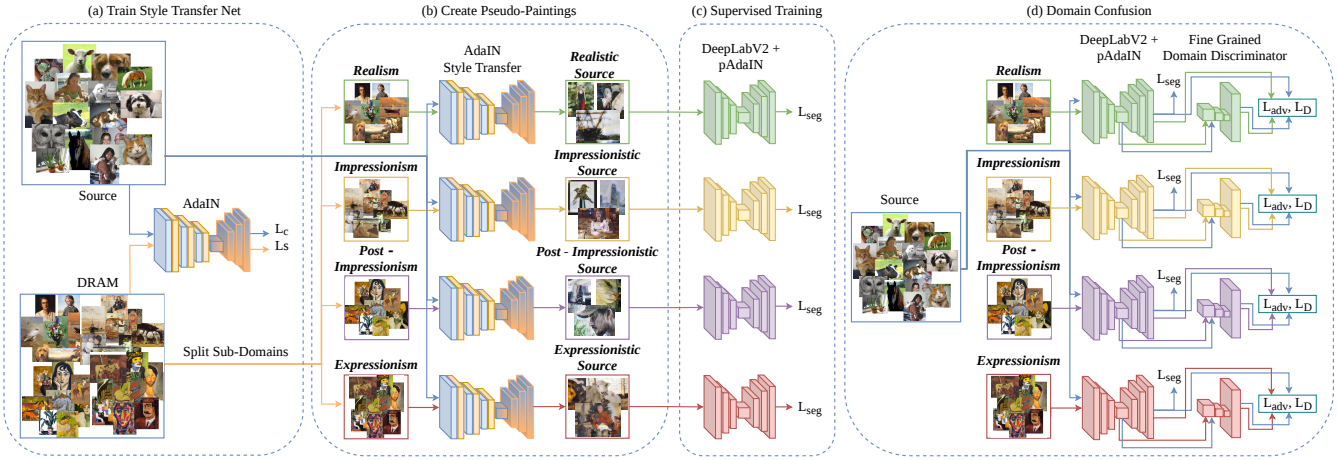

Figure 2: *Our proposed training flow: (a) we train a style transfer network using PASCAL VOC 2012 source data as content and DRAM training set as style images; (b) we split DRAM into four sub-domains and create pseudo-paintings for each by augmenting the original source dataset with each sub-domain as style images; (c) we train each of the domains separately in a supervised manner using DeepLabV2 [CBP*16] on the pseudo-paintings with their original labels from PASCAL VOC 2012; (d) we continue to refine the networks with a domain confusion step based on FADA [WSZ*20] and pAdaIN [NBW20]. We train with PASCAL VOC 2012 as source (without augmentations) and the DRAM sub domains as targets.*

work outputs a label per pixel in the image. Current state-of-the-art results on photographs of different datasets are achieved with DeepLabV3+ [CZP*18]. DeepLabV3+ uses a classification network as its encoder and up-samples the encoding output using atrous convolution layers followed by a simple convolutional decoder. We follow current domain adaptation methods and use DeepLabV2 [CBP*16] as the basic semantic segmentation network. DeepLabV2 does not use a convolutional decoder to up-sample its encodings. We compare our results to both DeepLabV2 and DeepLabV3+ as baselines. As recent papers such as [YCW19; LLC*21] have shown promising improvement on semantic segmentation by using modules based on self-attention transformers [VSP*17; DBK*20] we use HRNet [SZJ*19] with OCR [YCW19] as an additional baseline for comparison. As artistic paintings have differences in color, textures and geometric structure of objects, none of the three methods perform well in this domain when trained on real-life photographs.

A previous attempt to improve segmentation results on artistic paintings was made by Chatzistamatis et al. [CRT20] which focuses on recoloring art paintings to overcome color loss caused by color blindness by utilizing segmentation maps. To do so a pretrained MaskRCNN [HGDG17] is fine-tuned using annotated art paintings to output semantic maps which are used for the semantic recoloring process. Unlike Chatzistamatis et al. our work focuses on semantic segmentation of art in unrealistic challenging styles as well as over unseen art styles. Additionally, we present a unique unsupervised solution for semantic segmentation of art paintings and we compare our results to other recent methods.

Domain Adaptation (DA). Domain adaptation works with two datasets drawn from two different domains: a labeled dataset for the *source* domain and an unlabeled dataset for the *target* domain. The data of the target domain is the one we wish to optimize on a given task. Initial frameworks of domain adaptation targeted the image classification task and centered around adversarial domain confusion [GL15; THSD17; CWZ*20]. Later classification frameworks tackled more advanced challenged like multi-target domain adaptation [CZLL19; GSR*18] and some used artistic datasets [PBX*19; CWZ*20; GSR*18]. However, unlike our DRAM dataset the artistic datasets explored by these papers focused on the difference between different type of arts such as clip art and sketches rather than the subtle difference between fine art painting styles.

Initial semantic segmentation DA also used domain confusion [THS*18] and added an image translation module to reduce the gap between the source and target domains [HTP*18]. Later solutions mainly rely on three techniques: data augmentation, domain confusion, and self-learning. Each method uses these three techniques differently and may use only a subset of them. In the following, we elaborate on the first two techniques, data augmentation and domain confusion, as we do not use self-learning in our method. Additionally, we discuss common DA datasets and prior work related to our dataset challenges: target domain diversity and a large domain gap.

Datasets. As research in artistic domains is mainly focused around practical applications such as style transfer and art creation [GEB15; HB17; YNS19] most artistic datasets [Wik21a; KTH*14] do not come annotated for object identification or semantic segmentation. Other, more perceptual applications include artist/genre clustering [DTD*19] and art paintings classification [KTH*14]. Exceptions to this are [CZ14; PBX*19; LYSH17], but they include annotations only for image classification and not for semantic segmentation.

Current domain adaptation approaches focus on driving datasets where the source domain is 3D computer renderings and the target domain is realistic photographs. The most commonly used

are GTA5 [RVRK16] or Synthia [RSM*16] as the (synthetic) source domain, and Cityscapes [COR*16], BDD100K [YCW*20], or Cross-City [CCC*17] as the (realistic) target domain. In our paper we focus on the opposite transition, from realistic to synthetic domains, using artistic paintings as our target data. We created the DRAM dataset which focuses on semantic segmentation and presents a new and more general benchmark for the unsupervised domain adaptation task.

Data Augmentation. Augmentation is used to reduce the domain gap between the source data and the target data. Beyond simple geometric transformations, different image-level augmentations include CycleGAN [LYV19; YLSS20; HTP*18], Fourier-Domain window exchange [YS20] and Style Transfer as used by Banar et al. [BSG*21] to classify musical instruments in art paintings. Nuriel et al. [NBW20] utilize an AdaIN layer [HB17] randomly on different latent features to exchange information between the target and source data to reduce the pattern bias shown in [GRM*18] and [GEB15]. Li et al. [LYV19] and Yang el al. [YLSS20] also optimize the domain transformer between the framework learning steps to improve transformations by using information from the learning process.

Using CycleGAN [ZPIE17] as a transformation network from PASCAL VOC 2012 to DRAM resulted in unsatisfying results. We believe this may be due to the high complexity and diversity of the artistic domain. The DRAM training set contains over 50 different artists and we suspect that optimizing a single network for such a diverse domain may be too complex. We use AdaIN style transfer [HB17] for augmentation to reduce the domain gap between the realistic photographic source data and our artistic target data, by creating *pseudo-paintings*. We chose this method for its ability to preserve the overall look of the source image and for its speed, which enables augmenting large datasets relatively fast.

Domain Confusion. Domain Confusion is an adversarial method which utilizes a domain discriminator. The discriminator is used to train the semantic segmentation network's encoder to encode target and source samples to a joint domain, while optimizing the network over the labeled source data in a supervised fashion. In an effort to use the segmentation information extracted from the network, Tsai et al. [THS*18] uses two discriminators, one for encoder features and the other for output features. We follow FADA [WSZ*20] and pAdaIn [NBW20] which use a discriminator not only to distinguish between domains, but also to learn the class structure of the trained model to encourage a class-level alignment in the generated feature space.

Target Domain Diversity. Most domain adaptation methods assume that the target data is homogeneous. The setting where there are several different sub-domains comprising the target domain is known as Multi-Target Domain Adaptation [LMP*20; CCC*17; PWSK20; CZLL19; GSR*18; IJC*21]. In some, even the target domain labels are unknown [LMP*20; PWSK20; CZLL19]. We used painting labels as many datasets classify artistic paintings to their style based on expert knowledge [Wik21a]. In addition, we found that utilizing clustering of paintings in some feature space led to inferior segmentation results.

For Multi-Target Domain Adaptation, Liu et al. [LMP*20] use a curriculum training procedure where target images are used for training in order of their distance from the source domain. Park et al. [PWSK20] train a single segmentation network, but use a separate discriminator for each sub domain to separate their optimization process. Additionally, they perform a unique image transformation process for each sub-domain to better use the assumed separation of the sub domains. Isobe et al. [IJC*21] trains an "expert" segmentation network using AdaptSegNet [THS*18] for each sub domain and then trains another network using DA which uses the information learned by the expert networks.

As discussed in [IJC*21], a solution for a multi-target domain adaptation task, which on the one hand creates a single framework for flexible predictions of each sub-domain, and on the other hand achieves results that are as optimal as training with each domain separately, creates a challenging trade off. Our approach trains a separate segmentation network for each sub domain. Similarly to Park et al. [PWSK20] we train a unique image transformation network for each sub domain and apply it on our source data to reduce the domain gap. To create a flexible multi-domain framework we compose the networks predictions by the similarity of the input image style to the training images style, using the style representation presented by Gatys et al. [GEB15].

Domains Gap Domain Adaptation methods rely on a reasonable gap between the source and target domains to achieve good results. When the target domain is incoherent, reducing this gap becomes a nontrivial task as each target sub-domain may require a different approach. Current methods discussing this issue, such as [CCC*17; LMP*20], focus on class-level alignment, which helps align diverse sub-domains by class information rather than relying solely on global domain information. Dai et al. [DST*20] creates new "bridging" target domains which are trained on separate discriminators and help the network optimize as they are closer to the source domain, thus easing the large gap between the source and the target domain. Our method also utilizes class information in the domain confusion step. Furthermore, we show that for complex domains, understanding the unique properties of each sub-domain can help reduce the domain gap. As a result, we can achieve better results for each sub-domain separately and for the entire target dataset together, as well as for unseen artistic target domains.

3. DRAM Dataset

To our knowledge, the Diverse Realism in Art Movements (DRAM) dataset, is the first semantic segmentation dataset which uses artistic paintings as its target domain. The dataset is composed of 5677 unsegmented training images and 718 segmented test images of paintings of 152 different artists. The majority of the dataset (including all training images and 583 test images) is comprised of four diverse art movements: Realism, Impressionism, Post-Impressionism, and Expressionism.

The DRAM dataset was gathered mainly from the WikiArt art database [Wik21a] (5677 train images, 676 test images). The remaining images (42 test images) were gathered from Wikimedia [Com21] and Wikioo [Wik21b] image databases. The images were assigned to art movements using their original tags. To ensure we are learning an artistic movement rather than the style of a specific artist, no artist is shared between the training and the test sets of each movement.

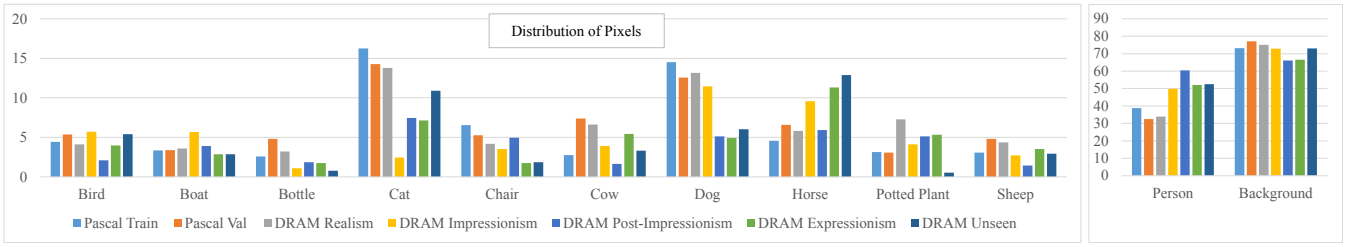

Figure 3: *The distribution of number of pixels per each class in PASCAL VOC 2012 training and validation sets compared to all movements in the DRAM test set. For the background class, we show the percentage of pixels labeled "Background" from all pixels. For all other classes, we present the percentage of pixels labeled per class excluding the background class to make the statistics more visible. Perfect equalization is difficult as some classes are more scarce in art than in real life photos.*

3.1. Domain Diversity

Four art movements were chosen for DRAM as they differ in terms of texture, geometry, and color. Their respective sizes are 1074, 1538, 1462 and 1603 images for training and 150, and 150, 142, 141 images for testing.

Realism: Emerged in France around the 1848 revolution. It sought to portray people from all classes of society in everyday situations with as much truth and accuracy as possible, without avoiding unpleasant aspects of life. For our purpose, this dataset is considered realistic in texture, geometry, and color; we consider it as the closest domain to our source dataset. To expand the Realism training set we added images from the Romanticism art movement as it shares realistic aspects with Realism in spite of it having more dramatic motifs.

Impressionism: Emerged in France in the 19th century. Characterized by thin yet visible brush strokes and an accurate depiction of light. This art movement is highly diverse in texture but is mostly realistic in geometry and color.

Post-Impressionism: Emerged in France roughly between 1886 and 1905 as a reaction against Impressionists' concern for realistic depiction of light and colors. Post-Impressionism emphasizes more abstract qualities and symbolic content. It has mild texture diversity, and is more diverse in geometry and colors.

Expressionism: Originated in Northern Europe around the beginning of the 20th century. Expressionist artists sought to express the meaning of emotional experience rather that physical reality. Expressionist paintings are highly diverse, using radically distorted geometry that represents a strong emotional effect. Expressionism does not oblige to any realistic concepts and is highly diverse in geometry and colors. It can also be diverse in texture, but not as strong as Impressionism.

The remaining 135 test images were gathered from eight art movements which do not appear in the training set: Art-Nouveau, Baroque, Cubism, Divisionism, Fauvism, Chinese Ink and Wash, Japonism and Rococo. As the training set does not include examples from these art movements, we consider these test images as unseen data to demonstrate the ability of our method to predict segmentation maps on data from unseen art movements that possibly present a larger domain gap and diversity.

Another aspect causing diversity in art movements is the choice of motifs. For example, the Baroque and Rococo movements are both fairly realistic styles, but Baroque style depicts elements from the Catholic Church with a strong religious atmosphere, while Rococo depicts reality in a more theatrical sense. Both movements appear in our unseen test set for validating generalization ability on such artistic variations.

3.2. Domains Gap

We use PASCAL VOC 2012 [EVW*12] as our source dataset with 11 classes common in art: Bird, Boat, Bottle, Cat, Chair, Cow, Dog, Horse, Sheep, Person, and Potted-Plant. The rest of the classes and any other object not recognized in our task are considered background, which is used as the 12th class in our tests. The test data in DRAM was manually segmented according to PASCAL VOC 2012 official guidelines with the only exception being labeling flower vases as potted-plant to expand the number of classes available (they share similar features and potted plants are scarce in most art movements).

To reduce the domain gap in terms of content and ensure a fair evaluation of our results, we took effort to equalize the class statistics between DRAM and PASCAL VOC 2012 (see Figure 3). Otherwise, any statistical measure could have been biased because of lack of a specific class or domination of another. We used an iterative process of adding paintings to the DRAM dataset while preserving similar distribution of classes.

3.3. Style Feature Space

The seminal work of Gatys et al. [GEB15] introduced the concept of Gram matrices, which are obtained by matrix multiplication of VGG19 convolution layers, and serve as an expression of the artistic texture-style of a given painting. We use this representation as a style feature space for our data. We concatenate the Gram representation of 5 pre-trained VGG19 layers: $conv11, conv21, conv31, conv41$ and $conv51$ as suggested by [GEB15]. For efficiency purposes, we use Kernal-PCA [TF09] with a cosine kernel to reduce the representation dimensions to 512, preserving more than 99% of the data variability. Figure 4 shows a TSNE plot [MH08] of the mapping of paintings from DRAM's four art movements compared to the mapping of PASCAL VOC 2012's photographs. The gap between the domains in terms of style is clear. To reduce this domain gap we use style transfer in the first step of our method as described in the next section.

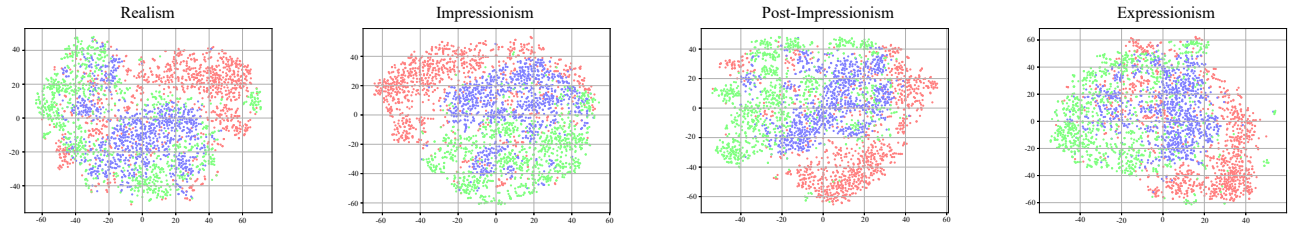

Figure 4: *TSNE plots of the style feature space of the four art movements in the DRAM dataset (green dots) compared to PASCAL VOC 2012 dataset (red dots) – the gap between each synthetic domain and the real domain is clear. From left to right we plot Realism, Impressionism, Post-Impressionism, and Expressionism. We use style-transfer to create a set of pseudo-paintings in each movement (blue dots) that bridge the gap between the domains, which are used for training in the first step of our method.*

4. Method

Given a target domain X_t, an image sampled from the target domain $x_t \in X_t$ and a finite set of classes $C = \{c_1, ...c_{|C|}\}$ we wish to assign the correct class to every pixel in x_t using the output of a chosen semantic segmentation network:

$$P_{x,y} = \arg\max_{c_i \in C} \phi_{x,y}(x_t) \tag{1}$$

where $\phi(\cdot)$ is the semantic segmentation model trained with domain adaptation and $\phi_{x,y}(x_t)$ is the vector of size $|C|$ taken from the (x,y) pixel location of the segmentation model output $P = \phi(x_t)$.

To train semantic segmentation on artistic paintings without excessive tagging we turn to an unsupervised domain adaptation framework, where the source domain (PASCAL VOC 2012) includes real photographs with ground-truth segmentation, and the target domain (DRAM) contains unsegmented paintings.

Most current domain adaptation methods also consider the target dataset samples as drawn from a single coherent domain. However, some image domains X_t may contain more than one sub-domain, and training it as a single domain achieves sub-optimal results due to the sub-domain differences. In our case, art paintings can differ considerably by many factors such as texture, color, and geometry. Treating many art movements as a single image domain produces a highly diverse domain which hurts the adaptation results.

To compensate for the large diversity, we suggest training each art movement as a single coherent sub-domain (see Figure 2). Doing so enables the trained sub-models to learn specific features relevant only to the specific art movement and prevents irrelevant features from misguiding the optimization. During inference, we combine these sub-models to segment not only paintings from the original domains, but also paintings from other, unseen art movements.

4.1. Augmentation and Pseudo-Paintings

The first step of our method utilizes style-transfer to create a dataset of segmented pseudo-paintings to use as a training set for a segmentation network. This step bridges the gap between the source and the target domain for training.

We first train a style transfer network using the style transfer approach suggested by Huang et al. [HB17]. This approach introduced the AdaIN layer that swaps statistics at the feature level from a style image to a content image we wish to stylize (Figure 2 (a)). One advantage of this method is that it can be trained on a collection of style images and then applied on arbitrary style and content images. We use the entire DRAM training set as style images and PASCAL VOC 2012 as content images (more details can be found in the supplemental materials).

Rather then using a single augmented dataset as common in current domain adaptation approaches, we augment the source dataset separately for each sub-domain. We use the original classification of the artwork and create four sub-domains in DRAM, namely: Realism, Impressionism, Post-Impressionism, and Expressionism. Each sub-domain is used separately as style images for augmenting the source domain and create a separate set of pseudo-paintings in the four different artistic styles (Figure 2 (b)). Figure 4 clearly shows how the new pseudo-painting datasets are closer in style feature space to the original paintings of each movement, and how the pseudo-paintings bridge the gap between the photographic domain of PASCAL VOC 2012 and the actual paintings. Examples of pseudo-paintings used for training our segmentation networks can be observed in Figure 5.

Using the pseudo-paintings of each sub-domain we train a semantic segmentation network for each sub-domain in a supervised fashion (Figure 2 (c)). As can be observed in Table 1, this training with augmented pseudo-paintings alone (marked as **OurMethod\DC**) improves the results of segmentation significantly for all sub-domain compared to the baseline without adaptation (**DeepLabV2**, **DeepLabV3+** and **HRNet+OCR**). This indicates that our style transfer augmentation helps reduce the domain gap between the source domain and each target sub-domain. However, we further adapt the segmentation networks using an additional domain confusion step as described next.

4.2. Domain Confusion

We chose to base our domain confusion step on FADA [WSZ*20] with pAdaIN [NBW20], since it achieves better regularization when trained on the source domain. This regularization is an important feature in the artistic domain because of its high diversity.

FADA is a domain adaptation framework which uses three training steps: supervised source training, domain confusion, and self-supervised learning. The domain confusion step uses a fine-grained domain discriminator which receives, in addition to the domain label, soft labels generated by the current network predictions. Doing

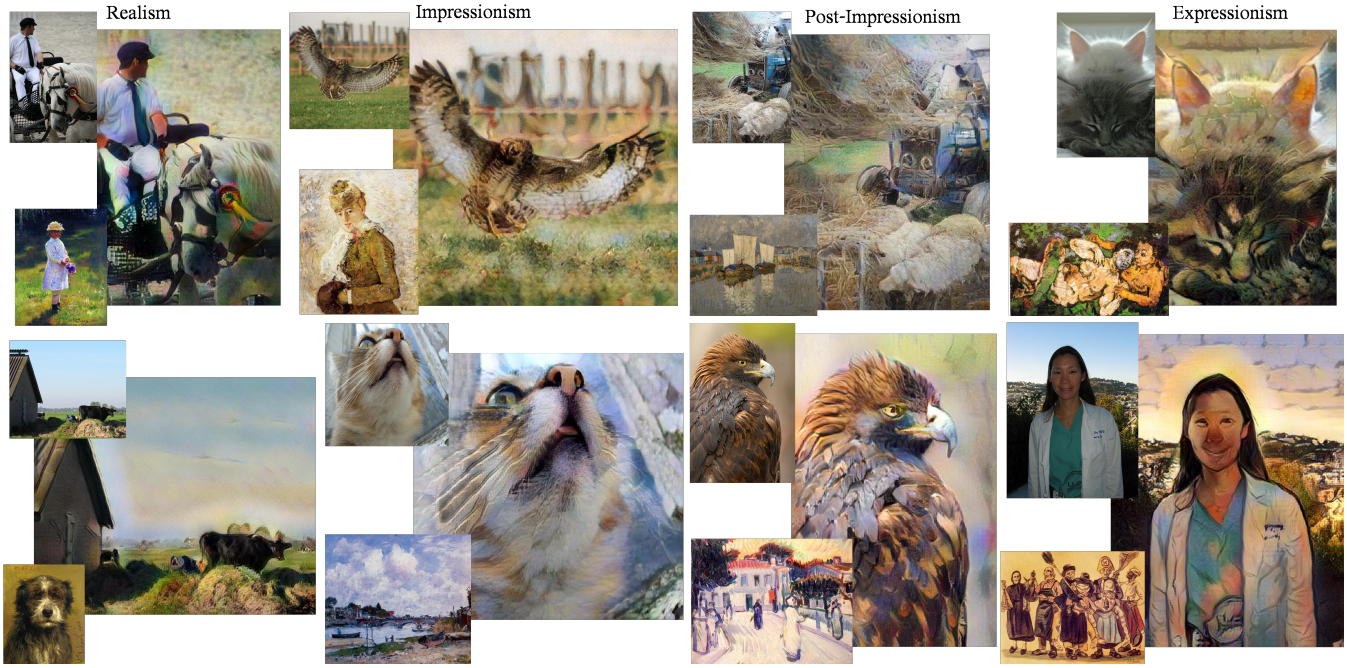

Figure 5: *Examples of pseudo-painting for the four art movements in the DRAM dataset from left to right: Realism, Impressionism, Post-Impressionism, and Expressionism. Each example shows the original photo and the original painting (left) used to create the pseudo-painting (right).*

so, the discriminator learns class information and achieves a better alignment between classes on the encoded latent domain. As discussed in section 2, this approach was found to be useful for highly diverse domains.

pAdaIN uses the framework proposed by FADA and presents a regularization term which reduces texture bias as presented by Geirhos et al. [GRM*18]. The method suggests adding an AdaIN layer [HB17] after every convolution layer in the ResNet encoder to swap image statistics between images among a batch with random probability of 0.01. The AdaIN layer swaps the image statistics while leaving global characteristics such as color and overall structure intact. Note that in the first and third steps of FADA, pAdaIN is used between random images inside a batch, where half of the batch is used as style images which transfer their statistics to the other half of the batch. In the second step, the target data batch is used as style images to transfer its statistics to the source data in the same fashion. More details regarding the training and networks can be found in the supplemental materials.

We have found that using self-supervised learning based on pseudo-labels (which is the third step of FADA) does not always improve the segmentation results on paintings. We believe this is due to the large diversity and gap between the photographic and artistic domains. Instead, our approach turns to combine the individual solutions of each sub-domain both for the known domains of our DRAM set, as well as for new artistic domains, as described next.

4.3. Multi-Domain Inference

Training each sub-domain separately introduces new challenges during inference. Art movements are based on abstract concepts rather than exact rules. It may be difficult to choose the correct sub-domain for every artistic painting, and especially for new ones from movements unseen during training. In addition, images tagged in a certain art movement can be closer to a different movement. Images may include artistic features from more than one art movement and can benefit from using the learned models of several art movements. Therefore instead of applying each sub-model separately, we combine them together using a method we term *multi-domain inference*.

As mentioned before, we use Gram matrices to represent images in style feature space. We pre-compute the Gram-representation of all training data, map them to the 512-dimensions feature space and store them. During inference, for each query image z, we search for its k-nearest neighbors in the style feature space and use the ratio of each sub domain as the weight for predictions of the different sub-domains. We define the weight w_z^i of each sub-domain i in the inference process as the percent of representatives from this sub-domain in the k-nearest neighbors of z:

$$w_z^i = \frac{1}{k}\sum_{i=1}^{k} 1_i(\theta_z^i) \tag{2}$$

where $1_i(\cdot)$ is the indicator function for the i'th sub-domain, and θ_z is the set of k-nearest neighbors of z.

If $\{\phi^i(\cdot)\}_{i=1}^n$ are the semantic segmentation models trained on the n sub-domains of X_t using domain adaptation, we use the

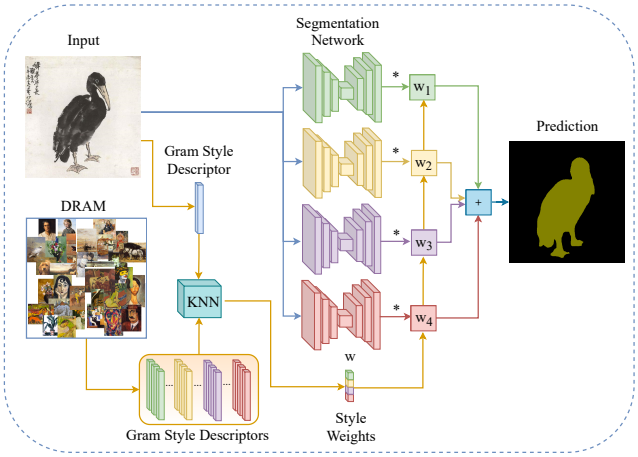

Figure 6: *Multi-Domain Inference pipeline. The input image is processed by each pre-trained art movement segmentation network. Then, a weight vector is calculated based on similarities of the image to the training set in Gram feature space and is used to combine the networks outputs and create a final segmentation prediction.*

weight vector w_z to combine the outputs of $\{\phi^i(\cdot)\}_{i=1}^n$ (see Figure 6). Thus, the flexible prediction of an unseen image z is defined for each pixel (x,y) as:

$$Q_{x,y} = \arg\max_{c \in C} \sum_{i=1}^{k} w_z^i \cdot \phi_{x,y}^i(z) \tag{3}$$

We use $k = 500$ for the KNN search. This k value allows greater precision, while smaller values tend to be more sensitive to outliers (e.g. using $k = 10$ results in a loss of up to 0.3% in accuracy).

Our experiments show that multi-domain inference can benefit results not only for new unseen art movements, but also for images of the original art movement sub-domains of the training set. This indicates that paintings from one art movement can contain features from more than one art movement.

5. Experiments

5.1. Implementation Details

Our experiments generally follow the hyper-parameters and augmentations strategy suggested by Wang et al. [WSZ*20]. In the following subsection, we elaborate on specific details and additional changes we made.

Datasets. As mentioned in Section 3, we use PASCAL VOC 2012 [EVW*12] as the source dataset and DRAM as our target dataset for domain adaptation of semantic segmentation on artistic paintings. PASCAL VOC 2012 is a semantic segmentation dataset containing real images annotated with 20 object classes and one background class. We use the dataset combined with the SBD dataset [HAB*11] as suggested by Chen et al. [CZP*18]. As we use only 12 classes of the original dataset, we filter out all images which do not contain at least one class of our 11 object classes, resulting in a total of 8362 source annotated images. To equalize the image resolutions, we resize all images in DRAM to have 500 pixels in their largest dimension while keeping original aspect ratio, similarly to PASCAL VOC 2012.

Style Transfer Augmentations. We train our style transfer network using PASCAL VOC 2012 as content images and DRAM train set as style images. We use this network to stylize PASCAL VOC 2012 separately for each of DRAM's four sub-domains training data. For each source image we use a single random style image with content/style weight parameter of 0.5. The network was trained on an Nvidia RTX2080Ti GPU for approximately 10 hours.

Domain Confusion Network. We use the FADA [WSZ*20] domain confusion framework with the enhancement of permuted AdaIN layers as presented by Nuriel et al. [NBW20]. Similar to previous domain adaptation methods, we use DeepLabV2 [CBP*16] with Resnet101 [HZRS16] backbone as the framework's semantic segmentation network. Since we do not use a validation set, we use the same settings used in [WSZ*20; NBW20] for GTA5 -> Cityscapes to train our networks. The only differences are that we use a batch size of 4 instead of 8 because of gpu memory limitations, and we resize images to 513×513 as suggested by Chen et al. [CZP*18] when training on PASCAL VOC 2012 dataset. As with our style transfer network, we trained the domain confusion step on an Nvidia RTX2080Ti GPU for approximately 10 hours.

Baselines and Benchmarks. Similarly to previous domain adaptation approaches, we use as baseline a DeepLabV2 network trained on PASCAL VOC 2012 and evaluate the results on our DRAM dataset. We also add the results of DeepLabV3+ and HRNet+OCR trained on PASCAL VOC 2012. We compare our adaptation results to the more classic domain adaptation framework AdaptSegNet [THS*18] and to three more recent domain adaptation frameworks: FDA [YS20], FADA [WSZ*20] and FADA+pAdain [NBW20]. We evaluate their results when trained with DRAM dataset as a unified target domain. Experimenting with artistic segmentation using multi-domain adaptation frameworks such as OCDA [LMP*20], and DHA [PWSK20] was more challenging as the implementation of these methods is missing. Instead, we used our method with the C-Driving dataset they use and report the results in the supplemental material. We use the mean intersection over union evaluation method (mIoU) for all experiments.

5.2. Results

Table 1 summarizes our results. As can be seen, over all sub-domains, as well as on unseen art styles, our method outperforms the alternative methods for semantic segmentation on artistic paintings. Table 2 breaks down the semantic segmentation results in art paintings by class and by artistic movement. As can be seen, the availability of class data in a certain art movement can heavily effect the results on a specific class, which in turn can bias the average. For this reason we took care to equalize class distributions.

Some qualitative results are shown in Figure 7. These demonstrate the improvement gained using our method for all four main art movements and four challenging unseen art movements: Divisionism, Fauvism, Ink & Wash and Rococo. More segmentation examples can be found in Section 6 and in the supplemental materials.

Method	Realism	Impressionism	Post-Impressionism	Expressionism	Unseen	DRAM
DeepLabV2	52.45	36.26	30.57	15.22	34.31	34.01
DeepLabV3+	60.02	39.27	29.40	15.49	30.59	35.17
HRNet+OCR	52.47	36.50	40.77	20.01	34.56	36.84
AdaptSegNet	45.25	35.55	35.34	21.06	36.57	34.60
FDA	39.89	31.89	32.08	17.87	22.66	29.84
FADA	61.00	42.19	44.23	24.57	38.15	43.10
FADA + pAdain	62.85	44.47	45.02	23.92	37.58	42.92
Our Method \DC	58.83	42.26	42.35	24.59	39.20	41.65
Our Method \ST	62.74	43.99	44.44	24.36	41.44	43.87
Our Method	63.41	45.99	47.28	27.37	42.03	45.71

Table 1: *Mean intersection over union (mIoU) for the test set of each sub-domain as well as for the whole DRAM dataset. We compare our method to four semantic segmentation domain adaptation methods, and to our method without the domain confusion step (\DC) and without style transfer (\ST). Please see details in Section 5.*

Domain	Background	Bird	Boat	Bottle	Cat	Chair	Cow	Dog	Horse	Person	P. Plant	Sheep	mIoU
DRAM	85.42	38.61	47.20	46.18	43.92	28.82	39.89	44.61	44.52	59.60	37.14	32.62	45.71
Realism	90.96	44.89	62.60	61.90	83.42	25.71	69.54	71.08	66.09	72.94	50.08	61.76	63.41
Impressionism	87.89	29.02	55.72	61.45	32.44	22.61	61.56	50.85	56.89	61.69	21.23	10.56	45.99
Post-Impressionism	84.26	36.69	46.81	57.95	44.24	41.47	21.00	52.51	52.76	67.94	46.80	14.92	47.28
Expressionism	77.93	28.45	20.70	29.63	10.55	16.38	15.10	17.08	32.91	43.05	26.23	10.49	27.37
Unseen	85.79	59.76	38.81	33.19	26.88	25.02	27.72	31.64	32.26	58.04	28.94	56.32	42.03

Table 2: *A breakdown of the mIoU for each class and each movement using our method for semantic segmentation in art paintings. The results can vary significantly also because of class occurrences in a specific movement.*

5.3. Ablation Study

First, as can be seen in Table 1, comparing our method with and without the style transfer components (see **OurMethod\DC\ST**) shows clearly that augmentation helps close the domain gap. Second, we can see that the domain confusion step improves the results on all artistic movements as well as the unseen ones compared to using only style-transfer augmentation (**OurMethod\DC\ST**).

We further study the effect of using our sub-domain training vs. training on the full dataset, and using different style-transfer settings for augmentation. We used style transfer with two settings: Using the entire DRAM training set and using each of its sub-domains separately to apply style transfer on the source data. We evaluate the effect of such style transfer when training DRAM as a single domain and when training sub-domain adaptation. The ablation study results can be found in Table 3.

As can be observed, using a unique augmented source for training each sub-domain proves to be more beneficial than using a single DRAM-augmented source. Specifically, using separate augmented source datasets with multi-domain adaptation achieves the highest improvement of 2.61% above previous approaches, 8.87% above the highest baseline, and 11.7% above the DeepLabV2 baseline for the entire DRAM test set. In addition, our approach improves each sub-domain result by up to 2.8% and is especially useful for the more challenging sub-domains. Another important aspect is the superiority of our approach on unseen art styles. Our results clearly show that our inference method achieves better adaptation for unseen art styles – all multi domain-inference results improve on previous approaches. Our method improves unseen data results by up to 3.88% above previous approaches, 7.47% above the highest baseline, and 7.72% above the DeepLabV2 baseline. This indicates that using style transfer and multi-domain adaptation helps achieve a more general model with better understanding of different art concepts and styles.

Another important aspect presented in Table 3 is that using the entire DRAM training set for augmentation is not effective. For both domain adaptation and multi-domain adaptation approaches using the entire DRAM training set for stylizing the source dataset causes a significant drop in the results. Because of the diversity of the artistic domain represented by DRAM, enabling the network to learn specific features for its different sub-domains results in a significant improvement.

6. Applications

Semantic segmentation is a fundamental task for understanding and using images in a wide range of applications. We demonstrate two applications of semantic segmentation that can be applied on fine art paintings. The first application focuses on analysis of artwork, and the second on synthesis.

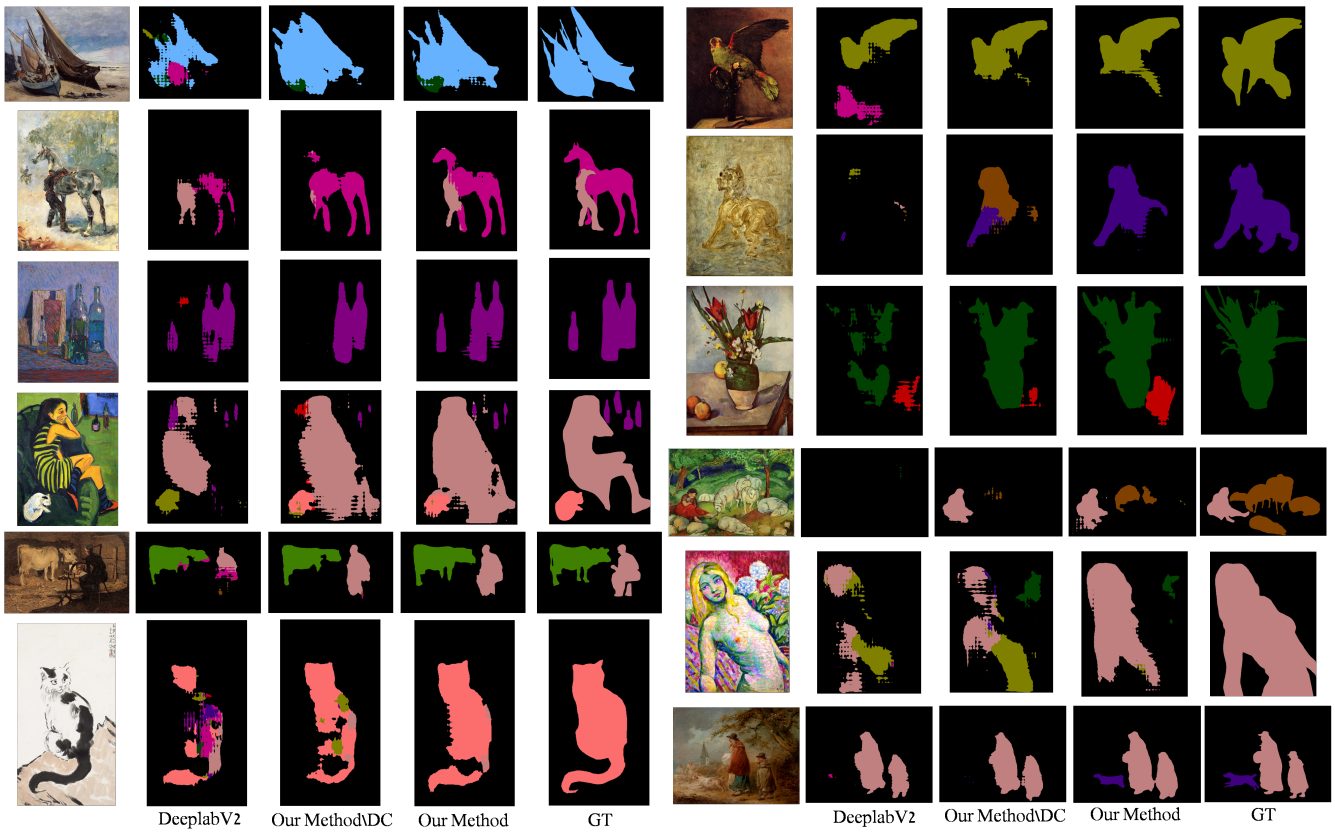

Figure 7: *Qualitative segmentation results on DRAM dataset. The leftmost image is the input image, and the rest (from left to right) present outputs of DeepLabV2 (baseline), Our method\DC (only the first step), Our full method, and ground-truth. Each row corresponds to a different art movement (from top to bottom): Realism, Impressionism, Post-Impressionism, and Expressionism. The last two rows showcase results from the unseen test set. From upper left to bottom right: Divisionism, Fauvism, Ink & Wash, and Rococo.*

Domains	Augemntation Style Data	Realism	Impressionism	Post Impressionism	Expressionism	Unseen	DRAM
Single Domain (DRAM)	No Augmentation	62.85	44.47	45.02	23.92	37.58	42.92
	DRAM	63.31	43.29	43.42	21.86	32.69	40.94
Multi Sub-Domains	No Augmentation	62.74	43.99	44.44	24.36	41.44	43.87
	DRAM	60.49	42.64	43.24	24.87	39.06	42.55
	Target Sub-Domain	63.41	45.99	47.28	27.37	42.03	45.71

Table 3: *Ablation study for our method using standard segmentation mIoU for evaluations. Results presented on training DRAM as a single domain (Single Domain) or training each sub-domain separately (Multi Sub-Domains). For each of those we evaluate training with no augmentations, source data augmented randomly by DRAM and source data augmented by each sub-domain separately (Target Sub-Domains).*

Figure 8: *A collection of dog segments from DRAM test set using our method. Such collections allow comparative analysis of artworks.*

6.1. Comparative Collections

To better understand and analyze artworks in a specific artistic movement or of a specific artist, comparisons are often performed between different paintings. Using semantic segmentation, such comparisons can be done not only at the painting level but also on specific objects or items. Using semantic segmentation one can gather all occurrences of a certain class from a given set of paintings, extract them from their original images and place them side-by-side for comparison. Figure 8 shows an example of gathering dogs from a set of paintings in the DRAM dataset. To create such a collection, we simply apply our method per painting and create the semantic maps. We then apply connected components labeling [Hor86] over all semantic maps. Lastly, we search for images that contain the specific class (Dog) and cut the relevant part out of the original painting. Such collections can also be used to detect and analyze segmentation errors. On a more abstract level, they depict the perception of a specific concept (Dog) by the network. More collections can be found in the supplemental materials.

6.2. Semantic Guided Style Transfer

Style-transfer, where a photograph is turned into a stylized image based on a chosen stylistic image, has become popular since its introduction in the work of Gatys et al. [GEB15]. Most style-transfer methods apply stylization on the whole image (e.g. Gatys et al. optimize the style loss over the entire image). However, it may be desirable to break the image to regions, recognize objects and apply a different stylization based on the semantics of the image's content.

Champandard [Cha16] presents a method for semantic style transfer, which requires input of two images (style and content) and two semantic maps (one for each image). The semantic style transfer method encodes the given images to Gram matrices and divides each Gram matrix into patches to find a nearest neighbor style image patch for every content image patch as suggested by Li et al. [LW16]. In addition Champandard uses semantic map encodings to add a semantic property to the nearest neighbors patch matching. This encourages the network to use patches taken from the corresponding semantic areas in the optimization process, resulting in a more accurate stylization per specific regions or objects (e.g. person->person, horse->horse, see Figure 9).

In its current form, the semantic style transfer method requires manual semantic maps, which can be challenging to create. Using our semantic segmentation method we automate the process and create an end-to-end framework for semantic style transfer that requires only the content/style pair of images. We use our method to create the semantic segmentation map of the style image, and the baseline method for the content image. These maps are used to find common classes and automatically match regions for the semantic style transfer method.

Figure 9 presents some example results. As can be observed, the specific style for each common class creates more coherent results, even when the semantic segmentation maps are not perfect.

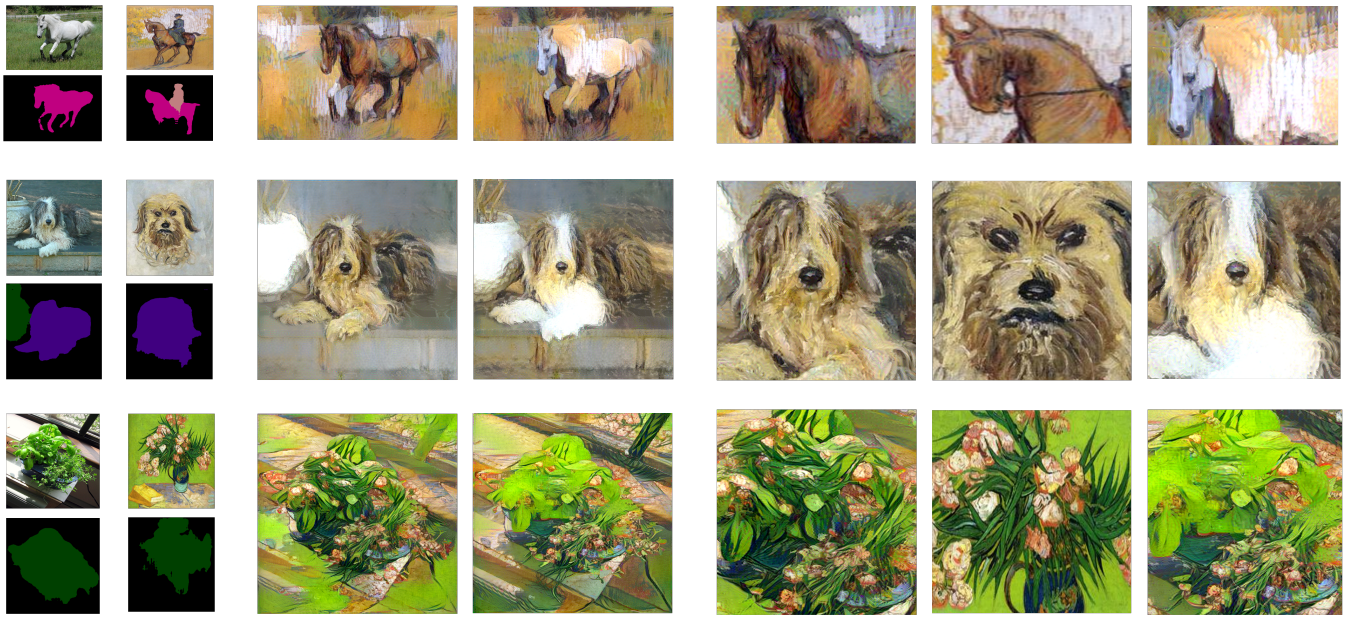

Figure 9: *Semantic style transfer examples. The content images were taken from PASCAL VOC 2012 dataset and the style images from the DRAM dataset. (a) shows the automatically segmented network inputs and their corresponding segmentation outputs. (b) shows the result of semantic style transfer and (c) the results of Gatys et al. [GEB15]. A closer look at specific semantic regions is shown next: (d) is a specific region in the stylized output and (e) is the corresponding semantic region in the style image. Comparison of (d) and (f) shows the same region without semantic stylization.*

Figure 10: *Examples of results of the Cubism movement (unseen dataset). Although there seems to be a hint of recognition (class recognition on the left and object localization on the right) results are unsatisfactory. We believe that achieving artistic geometric comprehension holds the solution for such geometrically challenging art domains.*

7. Discussion

Semantic segmentation is a difficult challenge in general, and more so in the artistic domain. Although our approach provides state-of-the-art results on artistic paintings, there is still a large gap to the ground truth segmentation, and room for improvement in future works. Note that a gap also exists in the results of segmentation of real photographs, although it is smaller. In the artistic domain the challenge is greater because of stylization. Our method addresses this by using the Gram-matrices based style feature space. Still, Gram-matrices have a bias towards color and texture. For example, in Figure 9, columns (c) and (f), optimizing the Gram representation of the image without semantic guidance results in output regions which preserve color but fewer brush strokes.

Painting also involves geometric stylizations. In many art movements such as Cubism, Expressionism, and Surrealism, the content of the image may be presented in a geometrically distorted fashion. Human perception can easily understand such paintings and their geometric structure, but this remains a challenge for computer algorithms. Previous methods such as Yaniv et al. [YNS19] achieved better performance by applying geometric augmentation (applying Affine transformations on the images). We experimented with a variety of geometric augmentation techniques and found that they can assist the results in more stylistic movements such as Post-Impressionism and Expressionism (providing an mIoU of 48.0, 27.92 instead of 47.28, 27.37, respectively). However, since we did not observe an absolute improvement in all artistic movements we decided not to apply geometric stylization by default in our method, and to leave this aspect for future work.

7.1. Limitations

Some of our segmentation outputs suffer from artifacts that appear as strokes of circles (see Figure 10). The reason for this effect is the up-sampling manner of the low dimensional prediction output of DeepLabV2. This happens as small prediction mistakes are exaggerated in the up-sampling process. Another effect caused by up-sampling is missing fine details around edges of ob-

jects. Small details disappear when down-sampled in the network, and are thus not considered for prediction. To solve these effects one can use larger images, but these may not always be available. DeepLabV3+ [CZP*18] offers a decoder module which decreases such effects, and using it along with domain adaptation may provide better results in the future. Additionally, as HRNet+OCR [SZJ*19; YCW19] generalizes better than both DeepLab models over DRAMs unrealistic movements when trained on PASCAL VOC 2012, using it with domain adaptation may also provide an improvement in future results.

Our method holds a few limitations which are derived from the above discussions. Since we do not consider geometric aspects to train our method, more abstract art movements have a smaller success rate than realistic ones. For more abstract movements such as Cubism, this can result in unsatisfactory results (Figure 10). Our method is based on prior knowledge of the movements of the training images to split them to sub-domains. In reality, art data may not hold such information. It may be possible in the future to use style feature spaces to automatically divide an input dataset to sub-domains.

Our dataset holds some limitation related to the complexity of gathering and annotating an art paintings segmentation dataset. As many art movements originated around the 19th century, they do not include modern classes like vehicles and electronic devices which are more common in photographs. This leads our dataset to settle for a relatively small number of classes. Another limitation resulting from annotation complexity is that our dataset lacks a pre-defined validation set. In the future we hope to expand DRAM's class coverage and gather more data as validation.

7.2. Conclusion

We presented a first semantic segmentation solution for artistic paintings. Our unsupervised approach for handling artistic domains achieves state of the art results in comparison to baseline methods for segmentation as well as alternative unsupervised domain adaptation methods. We presented the DRAM dataset that includes diverse examples of figurative art paintings and presents a new challenge for domain adaptation because of large domain gaps and large target domain diversity. We also showed that current domain adaptation approaches that focus on synthetic to realistic driving benchmarks, do not produce the same quality results when trained on realistic to synthetic benchmarks such as ours.

We defined a composite domain adaptation method that combines sub-domain solutions. We believe this flexible approach can be applied to different kinds of data domains by using better suited augmentation networks and by using different domain adaptation components.

While art creation applications have developed vastly in recent years, the field of art perception has been less explored. Our work takes a small step towards assisting the art perception of computers. We believe that more abstract geometry comprehension is a challenging aspect which may be the key for future advancement.

Acknowledgements

This research was partly supported by the Israel Science Foundation (grant No. 1390/19) and The Ministry of Innovation, Science and Technology (grant No. 16470-3).

References

[BSG*21] BANAR, NIKOLAY, SABATELLI, MATTHIA, GEURTS, PIERRE, et al. "Transfer Learning with Style Transfer between the Photorealistic and Artistic Domain". *Society for Imaging Science and Technology*. 2021 4.

[CBP*16] CHEN, LIANG-CHIEH, BARRON, JONATHAN T, PAPANDREOU, GEORGE, et al. "Semantic Image Segmentation with Task-Specific Edge Detection Using CNNs and a Discriminatively Trained Domain Transform". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 2, 3, 8.

[CCC*17] CHEN, YI-HSIN, CHEN, WEI-YU, CHEN, YU-TING, et al. "No More Discrimination: Cross City Adaptation of Road Scene Segmenters". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 4.

[Cha16] CHAMPANDARD, ALEX J. *Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artwork*. 2016. arXiv: 1603.01768 11.

[Com21] COMMONS, WIKIMEDIA. *Wikimedia Commons*. 2021. URL: https://commons.wikimedia.org/wiki/ 4.

[COR*16] CORDTS, MARIUS, OMRAN, MOHAMED, RAMOS, SEBASTIAN, et al. *The Cityscapes Dataset for Semantic Urban Scene Understanding*. 2016. arXiv: 1604.01685 2, 4.

[CRT20] CHATZISTAMATIS, STAMATIS, RIGOS, ANASTASIOS, and TSEKOURAS, GEORGE. "Image Recoloring of Art Paintings for the Color Blind Guided by Semantic Segmentation". May 2020, 261–273. ISBN: 978-3-030-48790-4 3.

[CWZ*20] CUI, SHUHAO, WANG, SHUHUI, ZHUO, JUNBAO, et al. *Gradually Vanishing Bridge for Adversarial Domain Adaptation*. 2020. arXiv: 2003.13183 [cs.CV] 3.

[CZ14] CROWLEY, ELLIOT J. and ZISSERMAN, ANDREW. "The State of the Art: Object Retrieval in Paintings using Discriminative Regions". *British Machine Vision Conference*. 2014 3.

[CZLL19] CHEN, ZILIANG, ZHUANG, JINGYU, LIANG, XIAODAN, and LIN, LIANG. *Blending-target Domain Adaptation by Adversarial Meta-Adaptation Networks*. 2019. arXiv: 1907.03389 [cs.LG] 3, 4.

[CZP*18] CHEN, LIANG-CHIEH, ZHU, YUKUN, PAPANDREOU, GEORGE, et al. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation". *The European Conference on Computer Vision (ECCV)*. 2018 2, 3, 8, 13.

[DBK*20] DOSOVITSKIY, ALEXEY, BEYER, LUCAS, KOLESNIKOV, ALEXANDER, et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". *CoRR* abs/2010.11929 (2020). arXiv: 2010.11929 3.

[DST*20] DAI, SHUYANG, SOHN, KIHYUK, TSAI, YI-HSUAN, et al. *Adaptation Across Extreme Variations using Unlabeled Domain Bridges*. 2020. arXiv: 1906.02238 [cs.CV] 4.

[DTD*19] DENG, YINGYING, TANG, FAN, DONG, WEIMING, et al. "Selective clustering for representative paintings selection". *Multimedia Tools and Applications* 78 (July 2019) 3.

[EVW*12] EVERINGHAM, M., VAN GOOL, L., WILLIAMS, C. K. I., et al. *The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results*. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html. 2012 2, 5, 8.

[GEB15] GATYS, LEON A., ECKER, ALEXANDER S., and BETHGE, MATTHIAS. *A Neural Algorithm of Artistic Style*. 2015. arXiv: 1508.06576 2–5, 11, 12.

[GL15] GANIN, YAROSLAV and LEMPITSKY, V. "Unsupervised Domain Adaptation by Backpropagation". *Proceedings of the 32nd International Conference on International Conference on Machine Learning (ICML)*. Vol. 37. 2015, 1180–1189 3.

[GRM*18] GEIRHOS, ROBERT, RUBISCH, PATRICIA, MICHAELIS, CLAUDIO, et al. *ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness*. 2018. arXiv: 1811.12231 4, 7.

[GSR*18] GHOLAMI, BEHNAM, SAHU, PRITISH, RUDOVIC, OGNJEN, et al. *Unsupervised Multi-Target Domain Adaptation: An Information Theoretic Approach*. 2018. arXiv: 1810.11547 [cs.CV] 3, 4.

[HAB*11] HARIHARAN, BHARATH, ARBELAEZ, PABLO, BOURDEV, LUBOMIR, et al. "Semantic contours from inverse detectors". *IEEE International Conference on Computer Vision (ICCV)*. 2011 8.

[HB17] HUANG, XUN and BELONGIE, SERGE. "Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization". *IEEE International Conference on Computer Vision (ICCV)*. 2017 3, 4, 6, 7.

[HGDG17] HE, KAIMING, GKIOXARI, GEORGIA, DOLLÁR, PIOTR, and GIRSHICK, ROSS B. "Mask R-CNN". *CoRR* abs/1703.06870 (2017). arXiv: 1703.06870 2, 3.

[Hor86] HORN, BERTHOLD K.P. *Robot Vision*. 1986 11.

[HTP*18] HOFFMAN, JUDY, TZENG, ERIC, PARK, TAESUNG, et al. "CyCADA: Cycle-Consistent Adversarial Domain Adaptation". *Proceedings of the 35th International Conference on Machine Learning*. Ed. by DY, JENNIFER and KRAUSE, ANDREAS. Vol. 80. Proceedings of Machine Learning Research. PMLR, July 2018, 1989–1998 3, 4.

[HZRS16] HE, KAIMING, ZHANG, XIANGYU, REN, SHAOQING, and SUN, JIAN. "Deep Residual Learning for Image Recognition". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 8.

[IJC*21] ISOBE, TAKASHI, JIA, XU, CHEN, SHUAIJUN, et al. *Multi-Target Domain Adaptation with Collaborative Consistency Learning*. 2021. arXiv: 2106.03418 [cs.CV] 4.

[KTH*14] KARAYEV, SERGEY, TRENTACOSTE, MATTHEW, HAN, HELEN, et al. "Recognizing Image Style". *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014 3.

[LLC*21] LIU, ZE, LIN, YUTONG, CAO, YUE, et al. "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows". *CoRR* abs/2103.14030 (2021). arXiv: 2103.14030 3.

[LMP*20] LIU, ZIWEI, MIAO, ZHONGQI, PAN, XINGANG, et al. "Open Compound Domain Adaptation". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020 4, 8.

[LW16] LI, CHUAN and WAND, MICHAEL. "Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis". *CoRR* abs/1601.04589 (2016). arXiv: 1601.04589 11.

[LWLW17] LUO, PING, WANG, GUANGRUN, LIN, LIANG, and WANG, XIAOGANG. "Deep dual learning for semantic image segmentation". *IEEE International Conference on Computer Vision (ICCV)*. 2017 2.

[LYSH17] LI, DA, YANG, YONGXIN, SONG, YI-ZHE, and HOSPEDALES, TIMOTHY M. *Deeper, Broader and Artier Domain Generalization*. 2017. arXiv: 1710.03077 [cs.CV] 3.

[LYV19] LI, YUNSHENG, YUAN, LU, and VASCONCELOS, NUNO. "Bidirectional learning for domain adaptation of semantic segmentation". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019 4.

[MH08] MAATEN, LAURENS VAN DER and HINTON, GEOFFREY. "Visualizing Data using t-SNE". *Journal of Machine Learning Research* 9.86 (2008), 2579–2605 5.

[NBW20] NURIEL, O., BENAIM, S., and WOLF, L. *Permuted AdaIn: Reducing the bias towards global statistics in image classification*. 2020. arXiv: 2010.05785 3, 4, 6, 8.

[PBX*19] PENG, XINGCHAO, BAI, QINXUN, XIA, XIDE, et al. "Moment matching for multi-source domain adaptation". *Proceedings of the IEEE International Conference on Computer Vision*. 2019, 1406–1415 3.

[PWSK20] PARK, KWANYONG, WOO, SANGHYUN, SHIN, INKYU, and KWEON, IN SO. "Discover, Hallucinate, and Adapt: Open Compound Domain Adaptation for Semantic Segmentation". *Advances in Neural Information Processing Systems (NIPS)*. Ed. by LAROCHELLE, H., RANZATO, M., HADSELL, R., et al. Vol. 33. Curran Associates, Inc., 2020, 10869–10880 4, 8.

[RPB15] RONNEBERGER, O., P.FISCHER, and BROX, T. "U-Net: Convolutional Networks for Biomedical Image Segmentation". *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Vol. 9351. LNCS. (available on arXiv:1505.04597 [cs.CV]). Springer, 2015, 234–241 2.

[RSM*16] ROS, GERMAN, SELLART, LAURA, MATERZYNSKA, JOANNA, et al. "The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes". *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 4.

[RVRK16] RICHTER, STEPHAN R., VINEET, VIBHAV, ROTH, STEFAN, and KOLTUN, VLADLEN. "Playing for Data: Ground Truth from Computer Games". *European Conference on Computer Vision (ECCV)*. Ed. by LEIBE, BASTIAN, MATAS, JIRI, SEBE, NICU, and WELLING, MAX. Vol. 9906. LNCS. Springer International Publishing, 2016, 102–118 2, 4.

[SZJ*19] SUN, KE, ZHAO, YANG, JIANG, BORUI, et al. *High-Resolution Representations for Labeling Pixels and Regions*. 2019. arXiv: 1904.04514 [cs.CV] 2, 3, 13.

[TF09] TREVOR HASTIE, ROBERT TIBSHIRANI and FRIEDMAN, JEROME. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.E. Second Edition.* Springer, 2009 5.

[THS*18] TSAI, Y.-H., HUNG, W.-C., SCHULTER, S., et al. "Learning to Adapt Structured Output Space for Semantic Segmentation". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018 3, 4, 8.

[THSD17] TZENG, E., HOFFMAN, J., SAENKO, K., and DARRELL, T. "Adversarial discriminative domain adaptation". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 3.

[VSP*17] VASWANI, ASHISH, SHAZEER, NOAM, PARMAR, NIKI, et al. *Attention Is All You Need*. 2017. arXiv: 1706.03762 [cs.CL] 3.

[Wik21a] WIKIART. *WikiArt Visual Art Encyclopedia*. 2021. URL: https://www.wikiart.org/ 3, 4.

[Wik21b] WIKIOO. *Wikioo.ord. Encyclopedia of Infinite Art*. 2021. URL: https://wikioo.org/ 4.

[WSZ*20] WANG, HAORAN, SHEN, TONG, ZHANG, WEI, et al. "Classes Matter: A Fine-grained Adversarial Approach to Cross-domain Semantic Segmentation". *The European Conference on Computer Vision (ECCV)*. 2020 3, 4, 6, 8.

[WZH*19] WU, HUIKAI, ZHANG, JUNGE, HUANG, KAIQI, et al. *FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation*. 2019. arXiv: 1903.11816 [cs.CV] 2.

[YCW*20] YU, FISHER, CHEN, HAOFENG, WANG, XIN, et al. "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020 4.

[YCW19] YUAN, YUHUI, CHEN, XILIN, and WANG, JINGDONG. "Object-Contextual Representations for Semantic Segmentation". *CoRR* abs/1909.11065 (2019). arXiv: 1909.11065 3, 13.

[YLSS20] YANG, YANCHAO, LAO, DONG, SUNDARAMOORTHI, GANESH, and SOATTO, STEFANO. "Phase Consistent Ecological Domain Adaptation". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020 4.

[YNS19] YANIV, J., NEWMAN, Y., and SHAMIR, A. "The face of art: Landmark detection and geometric style in portraits". *ACM Trans. Graph. Vol. 38 no. 4*. 2019, 60:1–60:15 3, 12.

[YS20] YANG, YANCHAO and SOATTO, STEFANO. "FDA: Fourier Domain Adaptation for Semantic Segmentation". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020 4, 8.

[ZPIE17] ZHU, JUN-YAN, PARK, TAESUNG, ISOLA, PHILLIP, and EFROS, ALEXEI A. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 4.

[ZSQ*17] ZHAO, HENGSHUANG, SHI, JIANPING, QI, XIAOJUAN, et al. "Pyramid Scene Parsing Network". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 2.

Appendix A: Appendix - List of Paintings

Figure 1 from top left:

Printemps. Charles Jacque. Wikimedia. Public Domain.
Fishing Boats, Calm Sea. Claude-Monet. Wikiart. Public Domain.
Grazing Horses. Franz Marc. Wikiart. Public Domain.
The Scream. Edvard Munch. Wikiart. Public Domain.
Cows in the Field. Constant Troyon. Wikiart. Public Domain.
Dun, a Gordon Setter Belonging to Comte Alphonse de Toulouse Lautrec. Henri de-Toulouse Lautrec. Wikiart. Public Domain.
At the Races. Henri de-Toulouse Lautrec. Wikiart. Public Domain.
Two Yellow Knots with Bunch of Flowers. Ernst Ludwig Kirchner. Wikiart. Public Domain.
Simpkin at the Tailor's Bedside. Beatrix Potter. Wikiart. Public domain.
The Green Line. Henri Matisse. Wikiart. Public Domain US.
Three Ducks. Xu Beihong. Wikiart. Public domain China.
Nature Morte Cubist. Louis Marcoussis. Wikiart. Public Domain.
Japanese. Vasily Vereshchagin. Wikiart. Public Domain.

Figure 2:

Realism Batch:
A Greenland, or Gyr Falcon. Archibald Thornburn. Wikiart. Public Domain.
A bouquet of flowers. Ilya Repin. Wikiart. Public Domain.
A Fisher Girl. Ilya Repin. Wikiart. Public Domain.
The Return from the Mill. Rosa Bonheur. Wikiart. Public Domain.
Brizo, a Shepherd's Dog. Rosa Bonheur. Wikiart. Public Domain.
Portrait of Lucy Langdon Williams Wilson. Thomas Eakins. Wikiart. Public Domain.
Gloucester Harbor. Winslow Homer. Wikiart. Public Domain.
Sheep on the Downs. James Ward. Wikiart. Public Domain.
Impressionism Batch:
Berck: Low Tide. Eugene Boudin. Wikiart. Public Domain.
Harnessed Horses. Eugene Boudin. Wikiart. Public Domain.
Mother and Child. John Henry Twachtman. Wikiart. Public Domain.
The red blouse. Berthe Morisot. Wikiart. Public Domain.
The Cage. Berthe Morisot. Wikiart. Public Domain.
Julie Manet and her Greyhound Laerte. Berthe Morisot. Wikiart. Public Domain.
On the Beach. Eduard Manet. Wikiart. Public Domain.
Lilac in a glass. Eduard-Manet. Wikiart. Public Domain.
Post-Impressionism Batch:
Head of Lorette with Curls. Henri Matisse. Wikiart. Public Domain.
Evening in a Russian Village. Konstantine Ivanovich Gorbatov. Wikiart. Public Domain.
Woman with necklace of gems. Pablo Picasso. Wikiart. Public Domain.
The picador. Pablo Picasso. Wikiart. Public Domain.
Arearea I. Paul Gauguin. Wikiart. Public Domain.
Self Portrait with mandolin. Paul Gauguin. Wikiart. Public Domain.
Aspidistra. Samuel Peploe. Wikiart. Public Domain.
Vase of Flowers. Eduard Vuillard. Wikiart. Public Domain.
Expressionism Batch:
Green Eye Mask. Amadeo De-Souza Cardoso. Wikiart. Public Domain.
Portrait of Francisco Cardoso. Amadeo De-Souza Cardoso. Wikiart. Public Domain.
The Greyhounds. Amadeo De-Souza Cardoso. Wikiart. Public Domain.
Girl. Sitting Female Nude. Max Pechstein. Wikiart. Public Domain.
The Masked Woman. Max Pechstein. Wikiart. Public Domain.
L'artiste et sa Femme. Gustave De-Smet. Wikiart. Public Domain.
La vie du Ferme. Gustave De-Smet. Wikiart. Public Domain.
Leopold Zborowski. Amedeo Modigliani. Wikiart. Public Domain.

Figure 5:

Girl with Flowers. Daughter of the Artist. Ilya Repin. Wikiart. Public Domain.
Winter (aka Woman with a Muff). Berthe Morisot. Wikiart. Public Domain.
Boats by the River Bank. Konstantin Gorbatov. Wikiart. Public Domain.
The Pagans. Oskar Kokoschka. Wikiart. Public Domain US.
Martin, a Terrier. Rosa Bonheur. Wikiart. Public Domain.
The Bridge over the Toques at Deauville. Eugene-Boudin. Wikiart. Public Domain.
Royan, Charente Inferieure. Samuel Peploe. Wikiart. Public Domain.
The Carnival. Paula Modersohn Becker. Wikiart. Public Domain.

Figure 6:

Cormorant. Xu Beihong. Wikiart. Public Domain China.

Figure 7

Fishing Boats on the Deauville Beach. Gustave Courbet. Wikiart. Public Domain.
The Green Parrot. Vincent Van-Gogh. Wikiart. Public Domain.
Artilleryman Saddling His Horse. Henri De-Toulouse Lautrec. Wikiart. Public Domain.
The Dog (Sketch of Touc). Henri De-Toulouse Lautrec. Wikiart. Public Domain.
Still Life with Bottles, Roderic O'Conor. Wikiart. Public Domain.
Still Life, Tulips and apples. Paul Cezanne. Wikiart. Public Domain.
Female Artist. Ernst Ludwig Kirchner. Wikiart. Public Domain.
Shepherdess with Sheep. Franz Marc. Wikioo. Public Domain.
Spinning. Giovanni Segantini. Wikiart. Public Domain.
Nu (Nude). Jean Metzinger. Wikiart. Public Domain US.
Cat. Xu Beihong. Wikiart. Public Domain China.
Woman, Child and Dog on a Road. George Morland. Wikiart. Public Domain.

Figure 9

Amazone. Henri De-Toulouse Lautrec. Wikiart. Public Domain.
Head of the Dog. Claude Monet. Wikiart. Public Domain.
Oleanders and Books. Vincent Van-Gogh, Wikiart. Public Domain.

Figure 10

Female nude (study). Pablo Picasso. Wikiart. Public Domain US.
The Tower of Blue Horses. Franz Marc. Wikiart. Public Domain.