Wasserstein Style Transfer

Youssef Mroueh
IBM Research & MIT-IBM Watson AI Lab

Abstract

We propose Gaussian optimal transport for image style transfer in an Encoder/Decoder framework. Optimal transport for Gaussian measures has closed forms Monge mappings from source to target distributions. Moreover, interpolating between a content and a style image can be seen as geodesics in the Wasserstein Geometry. Using this insight, we show how to mix different target styles, using Wasserstein barycenter of Gaussian measures. Since Gaussians are closed under Wasserstein barycenter, this allows us a simple style transfer and style mixing and interpolation. Moreover we show how mixing different styles can be achieved using other geodesic metrics between gaussians such as the Fisher Rao metric, while the transport of the content to the new interpolate style is still performed with Gaussian OT maps. Our simple methodology allows to generate new stylized content interpolating between many artistic styles. The metric used in the interpolation results in different stylizations. A demo is available on https://wasserstein-transfer.github.io.

1 Introduction

Image style transfer consists in the task of modifying an image in a way that preserves its content and matches the artistic style of a target image or a collection of images. Defining a loss function that captures this content/style constraint is challenging. A big progress in this field was made since the introduction of the neural style transfer in the seminal work of Gatys et al (Gatys et al., 2016, 2017). Gatys et al showed that by matching statistics of the spatial distribution of images in the feature space of deep convolutional neural networks (spatial Grammian), one could define a style loss function. In Gatys et al method, the image is updated via an optimization process to minimize this "network loss". One shortcoming of this approach is that is slow and that it requires an optimization per content and per style images. Many workarounds have been introduced to speedup this process via feedforward networks optimization that produce stylizations in a single forward pass (Johnson et al., 2016; Ulyanov et al., 2016; Li & Wand, 2016a; Wang et al., 2017). Nevertheless this approach was still limited to a single style image. (Ulyanov et al., 2017) introduced *Instance Normalization* (IN) to improve quality and diversity of stylization. Multiple styles neural transfer was then introduced in (Dumoulin et al., 2017) thanks to *Conditional Instance Normalization*(CIN). CIN adapts the normalized statistics of the transposed convolutional layers in the feedforward network with learned scaling and biases for each style image for a fixed number of style images. The concept of layer swap in (Chen & Schmidt, 2016) resulted in one of the first arbitrary style transfer. *Adaptive instance Normalization* was introduced in (Huang & Belongie, 2017) by making CIN scaling and biases learned functions from the style image, which enabled also arbitrary style transfer . The Whitening Coloring Transform (WCT) (Li et al., 2017a) which we discuss in details in Section 2 developed a simple framework for arbitrary style transfer using an Encoder/Decoder framework and operate a simple *normalization* transform (WCT) in the encoder feature space to perform the style transfer.

Our work is the closest to the WCT transform, where we start by noticing that instance normalization layers (IN,CIN, adaIN and WCT) are performing a transport map from the spatial distribution of a content

Proceedings of the 23[rd] International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

image to the one of a style image, via matching statistics of the distributions in a deep CNN feature space. First and second order statistics matching can be cast as Gaussian optimal transport. The Wasserstein geometry of Gaussian measures is very well studied in optimal transport (Takatsu, 2011) and Gaussian Optimal Transport (OT) maps have closed forms. We show in Section 3 that those normalization transforms are approximations of the OT maps. Linear interpolations of different content or styles at the level of those normalization feature transforms have been successfully applied in (Huang & Belongie, 2017; Dumoulin et al., 2017) we show in Section 4 that this can be interpreted and improved as Gaussian geodesics in the Wasserstein geometry . Furthermore using this insight, we show in Section 5 that we can define novel styles using Wasserstein barycenyter of Gaussians (Agueh & Carlier, 2011). We also extend this to other Fréchet means in order to study the impact of the ground metric used on the covariances in the novel style obtained via this non linear interpolation. Experiments are presented in Section 7.

2 Universal Style Transfer In the Lens of Optimal Transport

We review in this Section the approach of universal style transfer of WCT (Li et al., 2017a).

Encoding Map. Given a content image I_c and a style image I_s and a Feature extractor $F_j : \mathbb{R}^d \to \mathbb{R}^m, j = 1 \dots n$, where n is the spatial output of F, m is its feature dimension . Define the following *Encoding* map: $\boldsymbol{E} : \ I \in \mathbb{R}^d \to \nu_I = \frac{1}{n}\sum_{j=1}^{n} \delta_{F_j(I)} \in \mathscr{P}(\mathbb{R}^m)$ where $\mathscr{P}(\mathbb{R}^m)$ is the space of empirical measures on $\mathbb{R}^m$. For example F is a VGG (Simonyan & Zisserman, 2014) CNN that maps an image to $\mathbb{R}^{C\times(H\times W)}$($C$ is the number of channels, H the height and W the width). In other words the CNN defines a distribution in the space of dimension $m = C$, and we are given $n = H \times W$ samples of this distribution. We note $\nu = \mathbf{E}(\mathbf{I})$ this empirical distribution, i.e the spatial distribution of image I_c in the feature space of a deep convolutional network F.

Decoding Map. We assume that the encoding $\mathbf{E}$ is invertible , i.e exists: $\boldsymbol{D} : \ \nu \in \mathscr{P}(\mathbb{R}^m) \to \boldsymbol{D}(\nu) \in \mathbb{R}^d$, such that $\boldsymbol{D}(\boldsymbol{E}(I)) = I$. $(\boldsymbol{E}, \boldsymbol{D})$ is a VGG image Encoder/ Decoder for instance trained from the pixel domain to a spatial convolutional layer output in VGG and vice-versa.

Universal Style Transfer in Feature Space. Universal style transfer approach (Li et al., 2017a) works in the following way: WCT (Whitening Coloring Transform) defines a transform $\boldsymbol{T}_{c\to s}$ in the feature space $\mathbb{R}^m$: $\boldsymbol{T}_{c\to s} : \ x \in \mathbb{R}^m \to \boldsymbol{T}_{c\to s}(x) \in \mathbb{R}^m$, the style transfer Transform $\boldsymbol{T}_{c\to s}$ operates in the feature space and defines naturally a push forward map on the spatial distribution of the features of content image I_c:

$$\boldsymbol{T}_{c\to s,\#}(\nu(I_c)) := \frac{1}{n}\sum_{j=1}^{n} \delta_{\boldsymbol{T}_{c\to s}(F_j(I_c))}.$$

$\boldsymbol{T}_{c\to s}$ is defined so that the style transfer happens in the feature space i.e $\boldsymbol{T}_{c\to s,\#}(\nu(I_c)) = \nu(I_s)$. We obtain the stylized image $\tilde{I}_{c\to s}$ by decoding back to the image domain :

$$\tilde{I}_{c\to s} = \boldsymbol{D}(\boldsymbol{T}_{c\to s,\#}(\boldsymbol{E}(I_c))).$$

From this formalism we see that the universal style transfer problem amounts to finding a transport map $\boldsymbol{T}_{c\to s}$ from the spatial distribution of a content image in a feature space $\nu(I_c)$ to the spatial distribution of a target image in the same feature space $\nu(I_s)$. We show in the next section how to leverage optimal transport theory to define such maps. Moreover we show that the WCT transform and Adaptive instance normalization are approximations to the optimal transport maps.

3 Wasserstein Universal Style Transfer

Given $\nu_c = \boldsymbol{E}(I_c)$ and $\nu_s = \boldsymbol{E}(I_s)$, we formulate the style transfer problem as finding an optimal Monge map:

$$\inf_T \int \|x - T(x)\|_2^2 \, d\nu_c(x), \text{such that } T_{\#}(\nu_c) = \nu_s \tag{1}$$

the optimal value of this problem is $W_2^2(\nu_c, \nu_s)$, the Wasserstein two distance between ν_c and ν_s. Under some regularity conditions on the distributions, the optimal transport exists and is unique and $T_{c\to s}$ is the gradient of a convex potential (Benamou & Brenier, 2000).

Wasserstein Geometry of Gaussian Measures. Computationally Problem (1) can be solved using for example entropic regularization of the equivalent Kantorovich form of W_2^2 (Cuturi, 2013; Peyré & Cuturi, 2017) or in an end to end approach using automatic differentiation of a Sinkhorn loss (Frogner et al., 2015; Feydy et al., 2018) . We take here another route, and find the transport map that allows moment matching between the two distributions ν_c and ν_s. Using first and second moments of ν_c and ν_s, this amounts to computing the Wasserstein distance between the Gaussian approximation of ν_c and ν_s:

$$W_2^2(\mathcal{N}(m_{\mu_c},\Sigma_{\mu_c}),\mathcal{N}(m_{\nu_s},\Sigma_{\nu_s}))$$

where m_{μ_c},Σ_{μ_c} are means and covariance of ν_c, and m_{ν_s},Σ_{ν_s} of ν_s. The Wasserstein geometry of Gaussian measures is well studied and have many convenient computational properties (Takatsu, 2011), we summarize them in the following:

1) Closed Form W_2^2. Given two Gaussians distributions $\nu=\mathcal{N}(m_\mu,\Sigma_\mu)$, and $\mu=\mathcal{N}(m_\nu,\Sigma_\nu)$ we have: $W_2^2(\mathcal{N}(m_\mu,\Sigma_\mu),\mathcal{N}(m_\nu,\Sigma_\nu))=||m_\nu-m_\nu||^2+d^2_{\mathscr{B}}(\Sigma_\mu,\Sigma_\nu)$, where

$$d^2_{\mathscr{B}}(\Sigma_\mu,\Sigma_\nu)=\text{trace}\left(\Sigma_\mu+\Sigma_\nu-2\left(\Sigma_\mu^{\frac{1}{2}}\Sigma_\nu\Sigma_\mu^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)$$

is the Bures metric between covariances. The Bures metric is a goedesic metric on the PSD cone. (In Section 5.2 we discuss properties of this metric).

2) Closed Form Monge Map. The optimal transport map between two Gaussians with non degenerate covariances (full rank) has a closed-form: $T_{\mu\to\nu}:x\to m_\nu+A(x-m_\mu)$, where $A=\Sigma_\mu^{-\frac{1}{2}}\left(\Sigma_\mu^{\frac{1}{2}}\Sigma_\nu\Sigma_\mu^{\frac{1}{2}}\right)^{\frac{1}{2}}\Sigma_\mu^{-\frac{1}{2}}=A^\top$, i.e $T_{\mu\to\nu,\#}(\mu)=\nu$ and $T_{\mu\to\nu}$ is optimal in the W_2^2 sense. If the Gaussian were degenerate we can replace the square root matrices inverses with pseudo-inverses (Xia et al., 2014).

Gaussian Wasserstein Style Transfer. Using the Gaussian approximation we obtain a closed form optimal map from the content distribution to the style distribution as follows:

$$\boxed{\boldsymbol{T}^{\mathcal{W}}_{\nu_c\to\nu_s}(x)=\mu_s+A_{c\to s}(x-\mu_c),} \tag{2}$$

where $\mu_c=\frac{1}{n}\sum_{j=1}^n F_j(I_c),\mu_s=\frac{1}{n}\sum_{j=1}^n F_j(I_s)$, and $\Sigma_c=\frac{1}{n}\sum_{j=1}^n(F_j(I_c)-\mu_c)(F_j(I_c)-\mu_c)^\top$, and $\Sigma_s=\frac{1}{n}\sum_{j=1}^n(F_j(I_s)-\mu_s)(F_j(I_s)-\mu_s)^\top$ are means and covariances of ν_c and ν_s resp. and

$$A_{c\to s}=\Sigma_c^{-\frac{1}{2}}\left(\Sigma_c^{\frac{1}{2}}\Sigma_s\Sigma_c^{\frac{1}{2}}\right)^{\frac{1}{2}}\Sigma_c^{-\frac{1}{2}}.$$

Finally the Universal Wasserstein Style Transfer can be written in the following compact way, that is summarized in Figure 1:

$$\boxed{\tilde{I}_{c\to s}=\boldsymbol{D}(\boldsymbol{T}^{\mathcal{W}}{}_{\nu_c\to\nu_s,\#}(\boldsymbol{E}(I_c))).} \tag{3}$$

Relation to WCT and to Adaptive Instance Normalization. We consider two particular cases: *1) Commuting covariances and WCT (Li et al., 2017a).* Assuming that the covariances Σ_c and Σ_s commute meaning that $\Sigma_c\Sigma_s=\Sigma_s\Sigma_c$ (Σ_s and Σ_c have a common orthonormal basis) it is easy to see that the optimal transport map reduces to :

$$\boldsymbol{T}^{\mathcal{W}}_{\nu_c\to\nu_s}(x)=\mu_s+\Sigma_s^{\frac{1}{2}}\Sigma_c^{-\frac{1}{2}}(x-\mu_c)=\boldsymbol{T}^{\text{WCT}}_{c\to s}(x)$$

which is exactly the Whitening and Coloring Transform (WCT). Hence we see that WCT (Li et al., 2017a) is only optimal when the covariances commute (a particular case is diagonal covariances).

2) Diagonal Coviances and AdaIN, Instance Normalization (IN) and Conditional Instance Normalization (CIN)(Huang & Belongie, 2017; Ulyanov et al., 2017; Dumoulin et al., 2017). Let σ_s be the diagonal of Σ_s and σ_c be the diagonal of Σ_c. In case the covariances were diagonal it is easy to see that:

$$\boldsymbol{T}^{\mathcal{W}}_{\nu_c\to\nu_s}(x)=\mu_s+\sqrt{\sigma_s}\odot\frac{(x-\mu_c)}{\sqrt{\sigma_c}}=\text{AdaIN}(x),$$

this is exactly the expression of adaptive instance normalization AdaIN. We conclude that AdaIN, IN and CIN implement a diagonal approximation of the optimal Gaussian transport map ((μ_s,σ_s), are learned constant scaling and biases in IN and CIN , and are adaptive in adaIN).

4 Wasserstein Style/Content Interpolation with McCann Interpolates

One shortcoming of the formulation in problem (1) is that it does not allow to balance the content/style preservation as it is the case in end to end style transfer. Let $t\in[0,1]$ we formulate the style transfer problem with content preservation as follows:

$$\min_\nu(1-t)W_2^2(\nu,\nu_c)+tW_2^2(\nu,\nu_s), \tag{4}$$

The first term in Equation (4) measure the usual "content loss" in style transfer and the second term measures the "style loss". t balances the interpolation between the style and the content. In optimal transport theory, Problem (4) is known as the McCann interpolate (McCann, 1997) between ν_c and ν_s and the solution of (4) is a Wasserstein geodesic from ν_c to ν_s and is given by:

$$\nu_t=[(1-t)\text{Id}+tT_{\nu_c\to\nu_s}]_\#(\nu_c)$$

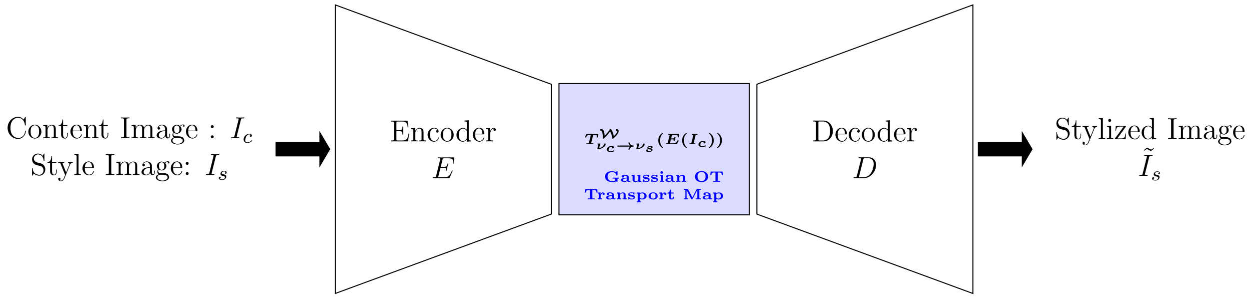

Figure 1: Wasserstein Style Transfer

Using again Gaussian optimal transport, we find a Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ that interpolates between the Gaussian approximations of source and target distributions as follows:

$$\min_{\nu \sim \mathcal{N}(\mu,\Sigma)} \mathscr{L}(\nu), \tag{5}$$

where $\mathscr{L}(\nu) = (1-t)W_2^2(\mathcal{N}(\mu,\Sigma), \mathcal{N}(\mu_c, \Sigma_c)) + tW_2^2(\mathcal{N}(\mu,\Sigma), \mathcal{N}(\mu_s, \Sigma_s))$. Fortunately this problem has also a closed form (McCann, 1997):

$$\nu_t = \mathcal{N}(\mu_t, \Sigma_t) = [(1-t)\mathrm{Id} + t\boldsymbol{T}^{\mathcal{W}}_{\nu_c \to \nu_s}]]_{\#}(\nu_c),$$

where $\boldsymbol{T}^{\mathcal{W}}_{\nu_c \to \nu_s}$ is given in Equation (2). $\{\nu_t\}_{t \in [0,1]}$ is a geodesic between ν_c and ν_s. Finally the Wasserstein Style/Content Interpolation can be written in the following compact way:

$$\nu_t = (1-t)\boldsymbol{E}(I_c) + t\boldsymbol{T}^{\mathcal{W}}{}_{\nu_c \to \nu_s, \#}(\boldsymbol{E}(I_c)), \tag{6}$$

$$\tilde{I}^t_{c \to s} = \boldsymbol{D}(\nu_t). \tag{7}$$

In practice both WCT and AdaIN propose similar interpolations in feature space, we give here a formal justification for this approach. This formalism allows us to generalize to multiple styles interpolation using the Gaussian Wasserstein barycenters.

5 Wasserstein Style Interpolation

Given $\{(I^j_s, \lambda_j)\}_{j=1\ldots S}$, S target styles images, and a content image (I_c, λ_{S+1}),where λ_j are interpolation factors such that $\sum_{j=1}^{S+1} \lambda_j = 1$. A naive approach to content/S styles interpolation can be given by:

$$\nu_\lambda = \sum_{j=1}^{S} \lambda_j \boldsymbol{T}^{\mathcal{W}}_{c \to s_j, \#}(\boldsymbol{E}(I_c)) + \lambda_{S+1}\boldsymbol{E}(I_c), \;\; I^\lambda_s = \boldsymbol{D}(\nu_\lambda),$$

this approach was proposed in both WCT and AdaIn by replacing $\boldsymbol{T}^{\mathcal{W}}$ by T^{WCT} and AdaIN respectively. We show here how to define a non linear interpolation that exploits the Wasserstein geometry of Gaussian measures.

5.1 Interpolation with Wasserstein Barycenters

Similarly to the content /style interpolation, we formulate the content / S styles interpolation problem as a Wasserstein Barycenter problem (Agueh & Carlier, 2011) as follows. Let $\nu^j_s = \boldsymbol{E}(I^j_s)$, and $\nu_c = \boldsymbol{E}(I_c)$, we propose to solve the following Wasserstein Barycenter problem:

$$\nu^s_\lambda = \arg\min_{\nu} \sum_{j=1}^{S} \lambda_j W_2^2(\nu, \nu^j_s) + \lambda_{S+1} W_2^2(\nu, \nu_c)$$

and then find the optimal map from ν_c to the barycenter measure ν^s_λ $T_{\nu_c \to \nu^s_\lambda}$. The final stylized image is obtained as follows: $\tilde{I}^\lambda_s = \boldsymbol{D}(\boldsymbol{T}_{\nu_c \to \nu^s_\lambda}(\boldsymbol{E}(I_c)))$.

Again we resort to Gaussian optimal transport, and solve instead the following problem:

$$\nu^s_\lambda = \arg\min_{\nu \sim \mathcal{N}(\mu,\Sigma)} \mathscr{L}(\nu), \tag{8}$$

where $\mathscr{L}(\nu) = \sum_{j=1}^{S} \lambda_j W_2^2(\mathcal{N}(\mu,\Sigma), \mathcal{N}(\mu^j_s, \Sigma^j_s)) + \lambda_{S+1} W_2^2(\mathcal{N}(\mu,\Sigma), \mathcal{N}(\mu_c, \Sigma_c))$. As shown by Agueh and Carlier (Agueh & Carlier, 2011) the Wasserstein Barycenter of Gaussians is itself a Gaussian $\nu^s_\lambda = \mathcal{N}(\bar{\mu}_\lambda, \bar{\Sigma}_\lambda)$, where $\bar{\mu}_\lambda = \sum_{j=1}^{S} \mu^s_j + \lambda_{S+1}\mu_c$, and $\bar{\Sigma}_\lambda$ is a Bures Mean. Noting $\Sigma^{S+1}_s = \Sigma_c$ we have:

$$\bar{\Sigma}_\lambda = \arg\min_{\Sigma} \sum_{j=1}^{S+1} \lambda_j d^2_{\mathscr{B}}(\Sigma, \Sigma^j_s)$$

Agueh and Carlier showed that $\bar{\Sigma}_\lambda$ is the unique positive definite matrix solution of the following fixed point problem: $\bar{\Sigma} = \sum_{j=1}^{S+1} \lambda_j \left(\bar{\Sigma}^{\frac{1}{2}} \Sigma^j_s \bar{\Sigma}^{\frac{1}{2}}\right)^{\frac{1}{2}}$. In order to solve this problem we use an alternative fixed point strategy proposed in (Álvarez Esteban et al.), since it

Figure 2: Wasserstein Barycenter Interpolation between a content image and two target styles images. The weights $\{\lambda_j\}$ used are given above the two examples.

converges faster in practice. For $\ell = 0 \dots L-1$:

$$\bar{\Sigma}_\ell = \sum_{j=1}^{S+1} \lambda_j \bar{\Sigma}_{\ell-1}^{-\frac{1}{2}} \left(\bar{\Sigma}_{\ell-1}^{\frac{1}{2}} \Sigma_s^j \bar{\Sigma}_{\ell-1}^{\frac{1}{2}} \right)^{\frac{1}{2}} \bar{\Sigma}_{\ell-1}^{-\frac{1}{2}}, \tag{9}$$

and we initialized as in (Xia et al., 2014): $\bar{\Sigma}_0 = \Sigma_s^{j_0}, j_0 = \arg\max_{j=1\dots S+1} \lambda_j$, we found that $L = 50$ was enough for convergence, i.e we set $\bar{\Sigma}_\lambda = \bar{\Sigma}_L$. Matrix square root and inverses were computed using SVD which gives an overall complexity of $O(Lm^3)$ and we used truncated SVD to stabilize the inverses. Finally since the Barycenter is a Gaussian , the optimal transport map from the Gaussian spatial content distribution $\mathcal{N}(\mu_c, \Sigma_c)$ to the barycenter (mix of styles and content) $\mathcal{N}(\bar{\mu}_\lambda, \bar{\Sigma}_\lambda)$ is given in closed form as in Equation (2):

$$\boxed{\boldsymbol{T}^{\mathscr{W}}_{\nu_c \to \nu^s_\lambda}(x) = \bar{\mu}_\lambda + \Sigma_c^{-\frac{1}{2}} \left(\Sigma_c^{\frac{1}{2}} \bar{\Sigma}_\lambda \Sigma_c^{\frac{1}{2}} \right)^{\frac{1}{2}} \Sigma_c^{-\frac{1}{2}} .(x - \mu_c)}. \tag{10}$$

Finally to obtain the stylized image as a result of targeting the mixed/style ν^s_λ we decode back:

$$\tilde{I}^\lambda_{cs} = \boldsymbol{D}(\boldsymbol{T}^{\mathscr{W}}_{\nu_c \to \nu^s_\lambda}(\nu_c)).$$

Figures 2 and 3 give an example of our approach for mixing content images with style images. We see that the Wasserstein barycenter captures not only the color distribution but also the details of the artistic style (for instance Frida Kahlo's unibrow is well captured smoothly in the transition between Picasso self portrait and Frida Kahlo).

Figure 3: Wasserstein barycenters for Style Mixing and Transfer. The content image on the right corner of the triangle is mixed with the two styles images. Each image in the triangle correspond to a set of interpolation weights defined by proximity to the content or style images in the triangle.

5.2 Style Interpolation with Fréchet Means

In the previous section we defined interpolations between the content and the styles images. In this section we define a "novel style" via an interpolation of style images only, we then map the content to the novel style using Gaussian optimal transport.

From Wasserstein Barycenter to Fréchet Means on the PSD manifold. As discussed earlier the Wasserstein Barycenter of the Gaussian approximations of the spatial distribution of style images in CNN feature spaces can be written as:

$$\min_{\mu,\Sigma} \sum_{j=1}^{S} \lambda_j \left(d^2_{\mu}(\mu, \mu^j_s) + d^2_{\text{cov}}(\Sigma, \Sigma^j_s) \right), \tag{11}$$

for $d^2_{\mu}(\mu, \mu') = \|\mu - \mu'\|^2_2$ the euclidean metric , $d^2_{cov}(\Sigma, \Sigma') = d^2_{\mathscr{B}}(\Sigma, \Sigma')$, the Bures metric.

The Bures Metric is a geodesic metric on the positive definite cone and and has another representation as a procrustes registration metric (Masarotto et al., 2018):

$$d^2_{\mathscr{B}}(\Sigma, \Sigma') = \min_{U \in \mathbb{R}^{m\times m}, UU^\top = I} \left\| \sqrt{\Sigma} - \sqrt{\Sigma'}U \right\|^2_F.$$

From this we see the advantage of Wasserstein barycenter on for example using $d^2_{\text{cov}}(\Sigma, \Sigma') = \|\Sigma - \Sigma'\|^2_F$ the Frobenius norm. Bures Metric aligns the the square root of covariances using a rotation. From this we see that by defining a new metric on covariances we can get different form of interpolates, we fix $d^2_{\mu}(\mu, \mu') = \|\mu - \mu'\|^2_2$, and hence on μ we use always the arithmetic mean $\mu_{arth} = \sum_{j=1}^{S} \lambda_j \mu^j_s$. We give here different metrics d^2_{cov} that defines different Fréchet means on the PSD manifold (see (Bhatia, 2013) and references there in):

1) Arithmetic Mean: Solving Eq. (11) for $d^2_{\text{cov}}(\Sigma, \Sigma') = \|\Sigma - \Sigma'\|^2_F$, we define the target style $\nu^{\lambda,\text{arth}}_s = \mathcal{N}(\mu_{arth}, \Sigma_{\text{arth}})$, where $\Sigma_{\text{arth}} = \sum_{j=1}^{S} \lambda_j \Sigma^s_j$.

2) Harmonic Mean: Solving Eq. (11) for $d^2_{\text{cov}}(\Sigma, \Sigma') = \left\|\Sigma^{-1} - \Sigma'^{-1}\right\|^2_F$, we define the target style $\nu^{\lambda,\text{Harm}}_s = \mathcal{N}(\mu_{\text{arth}}, \Sigma_{\text{Harm}})$, where $\Sigma_{\text{Harm}} = \left(\sum_{j=1}^{S} \lambda_j (\Sigma^s_j)^{-1} \right)^{-1}$.

3) Fisher Rao Mean (Karcher or Geometric Mean). For $d^2_{cov}(\Sigma, \Sigma') = \left\| \log(\Sigma^{-\frac{1}{2}} \Sigma' \Sigma^{-\frac{1}{2}}) \right\|^2_F = 2\mathscr{F}^2(\mathcal{N}(0, \Sigma), \mathcal{N}(0, \Sigma'))$, that is the Riemannian natural metric or the Fisher Rao metric between Centered Gaussians. log here refers to matrix logarithm. The Fisher Rao metric is a geodesic distance and its metric tensor is the Fisher information matrix . Solving Eq. (11) with the Fisher Rao metric we obtain the so called Karcher Mean between PSD matrices $\Sigma_{\text{FisherRao}}$, and we define the target style $\nu^{\lambda,\text{FisherRao}}_s = \mathcal{N}(\mu_{\text{arth}}, \Sigma_{\text{FisherRao}})$.
In order to find the Karcher mean we use manifold optimization techniques of (Zhang & Sra, 2016) as follows. The gradient manifold update is :

$$\Sigma_\ell = \Sigma^{\frac{1}{2}}_{\ell-1} \exp\left(-\eta \sum_{j=1}^{S} \log\left(\Sigma^{\frac{1}{2}}_{\ell-1} (\Sigma^s_j)^{-1} \Sigma^{\frac{1}{2}}_{\ell-1} \right) \right) \Sigma^{\frac{1}{2}}_{\ell-1}, \tag{12}$$

we initialize Σ_0 as in the Wasserstein case and iterate for $L = 50$ iterations with η the learning rate set to 0.01.

Remark 1. *While we defined here the barycenter style of each metric as a Gaussian, Wasserstein Barycenter is the only one that guarantees a Gaussian barycenter (Agueh & Carlier, 2011).*

Mapping a content image to a target novel style. Given now the new style $\nu^{\lambda,\text{mean}}_s$, where mean is in {arth, harm, Fisher Rao, Wasserstein}, we stylize a content image I_c using Gaussian Optimal transport as described in the paper:

$$\boxed{\tilde{I}_{c\to s} = \boldsymbol{D}(\boldsymbol{T}^{\mathcal{W}}_{\nu_c \to \nu^{\lambda,\text{mean}}_s, \#}(\boldsymbol{E}(I_c))).}$$

Our approach is summarized in Algorithms 1 and 2 .

Algorithm 1 FRECHET MEAN STYLE INTERPOLATION AND CONTENT STYLIZATION(d_{cov})

Inputs: $\{I^j_s\}_{j=1\dots S}$ style images, content Image I_c , interpolations weights $\{\lambda_j\}_{j=1\dots S+1}$, Encoder/Decoder $(\boldsymbol{E}, \boldsymbol{D})$,.
Encode: $\nu_c = \boldsymbol{E}(I_c), \nu^j_s = \boldsymbol{E}(I^j_s), j = 1 \dots S$
Statistics: $(\mu_c, \Sigma_c), (\mu^j_s, \Sigma^j_s), j = 1 \dots S$
Content/Style or Style only: if content/style $\mu^s_{S+1} = \mu_c, \Sigma^s_{S+1} = \Sigma_c, , S \leftarrow S + 1$, else pass.
Target Bary Mean: $\bar{\mu}_\lambda = \sum_{j=1}^{S} \lambda_j \mu_j$.
Target Bary Covariance: $\bar{\Sigma}_\lambda$ =FRECHET MEAN($\{\lambda_j, \Sigma^j_s\}, d_{cov}$)
Novel Style: $\nu^\lambda_s = \mathcal{N}(\bar{\mu}_\lambda, \bar{\Sigma}_\lambda)$
Gaussian OT Content to Target: Compute $\nu_{cs} = \boldsymbol{T}^{\mathscr{W}}_{\nu_c \to \nu^s_\lambda, \#}(\nu_c)$ given in Eq. (10)
Decode: $\boldsymbol{D}(\nu_{cs})$

Algorithm 2 FRECHET MEAN($\{\lambda_j, \Sigma_s^j\}, d_{\text{cov}}$)

Initialize: $\bar{\Sigma}_0 = \Sigma_s^{j_0}, j_0 = \arg\max_{j=1\ldots S} \lambda_j$
if $d_{cov} = d_{\textbf{Bures}}$ find $\bar{\Sigma}_\lambda$ solve using iterations in Eq (9)
if $d_{cov} = d_{\textbf{Fisher Rao}}$ find $\bar{\Sigma}_\lambda$ solve using iterations in Eq (12)
if $d_{cov} = d_{\textbf{Frobeinus}}$ $\bar{\Sigma}_\lambda = \sum_j \lambda_j \Sigma_j^s$
if $d_{cov} = d_{\textbf{Harmonic}}$ $\bar{\Sigma}_\lambda = (\sum_j \lambda_j (\Sigma_j^s)^{-1})^{-1}$

6 Related works

OT for style Transfer and Image coloring. Color transfer between images using regularized optimal transport on the color distribution of images (RGB for example) was studied and applied in (Ferradans et al., 2013). The color distribution is not gaussian and hence the OT problem has to be solved using regularization. Optimal transport for style transfer using the spatial distribution in the feature space of a deep CNN was also explored in (Marron, 2018; Kolkin et al., 2019; Lu et al., 2019). Marron (2018) uses W_2^2 for Gaussians as content and style loss and optimizes it in an end to end fashion similar to (Gatys et al., 2016, 2017). (Kolkin et al., 2019) uses an approximation of the Wasserstein distance as a loss that is also optimized in an end to end fashion. Both approaches don't allow universal style transfer and an optimization is needed for every style/content image pairs. Lu et al. (2019) is closely related to our work and uses a form of Gaussian optimal transport but does not extend to the interpolation case using OT barycenters.

Wasserstein Barycenter for Texture Mixing. Similar to our approach for Wasserstein mixing in an encoder/decoder framework, (Rabin et al., 2012) uses the wavelet transform to encode textures, applies Wasserstein barycenter on wavelets coefficients, and then decodes back using the inverse wavelet transform to synthesize a novel mixed texture. The wasserstein barycenter problem there has to be solved exactly and the Gaussian approximation can not be used since the wavelet coffecient distribution is not Gaussian. A special model for Gaussian texture mixing was developed in (Xia et al., 2014) and (Wang et al., 2018). Interpolating different styles was also addressed in (Wynen et al., 2018) based on archetypal analysis.

Other approaches to style Transfer. While our focus in this paper was on OT metrics for style transfer other approaches exist (see (Jing et al., 2017) for a review) and have used different type of losses such as MRF loss (Li & Wand, 2016b) , MMD loss (Li et al., 2017b), GAN loss (Li & Wand, 2016a) and cycle GAN loss (Zhu et al., 2017).

7 Experiments

In order to test our approach of geometric mixing of styles we use the WCT framework (Li et al., 2017a), where we use a pyramid of 5 encoders $(E_r, D_r), r = 1\ldots 5$ at different spatial resolutions, where (E_5, D_5) corresponds to the coarser resolution, and (E_1, D_1) the finer resolution. Following WCT we use a coarse to fine approach to style transfer as follows. Given interpolation weights $\{\lambda_j, j = 1\ldots S\}$, we start with $r = 5$ and with $\nu_c = E_5(I_c)$:

1. We encode all style images at resolution r, $E_r(I_s^j), j = 1\ldots s$. We define the mixed style $\nu_s^{\lambda,r}$ at resolution r using one of the mixing strategies (Frechet Mean) in Section 4, using Algorithms 1 and 2.

2. We find the Wasserstein Transport map at resolution r between the content ν_c and the novel style $\nu_s^{\lambda,r}$ and compute the transformed features: $\nu_{cs}^r = \boldsymbol{T}^{\mathcal{W}}_{\nu_c \to \nu_s^{\lambda,r},\#}(\nu_c)$.

3. We decode the novel image at resolution $r : I_c^r = D_r(\nu_{cs}^r)$.

4. We set $\nu_c = E_{r-1}(I_c^r)$, then set r to $r-1$ and go to step 1 until reaching $r = 1$.

The stylized output of this procedure is I_c^1. We also experimented in the Appendix with applying the same approach but in fine to coarse way starting from the higher resolution $r = 1$ to the lower resolution encoder $r = 5$. We show in Figure 4 the output of our mixing strategy using two of the geodesic metrics namely Wasserstein and Fisher Rao barycenters. We give as baseline the AdaIn output for this (this same example was given in (Huang & Belongie, 2017) we reproduce it using their available code). We show that using geodesic metrics to define the mixed style successfully capture the subtle details of different styles. More examples and comparison to literature and other types of mixing can be found in the Appendix.

8 Conclusion

We showed in this paper that universal style transform, that performs moment matching of source and

Table 1: Wasserstein Barycenter Mixing and Wasserstein Transport

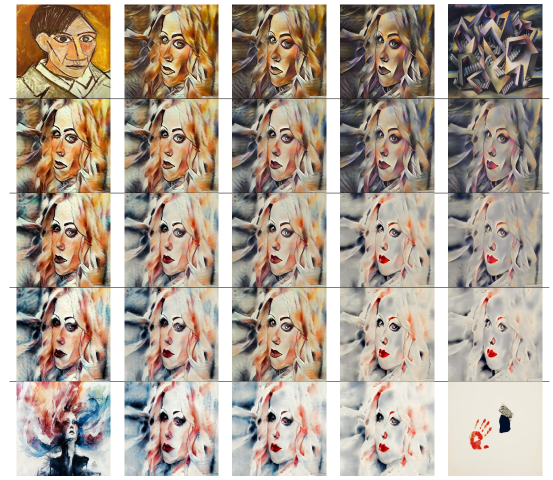

Table 2: Karcher (Fisher Rao) Barycenter Mixing and Wasserstein Transport

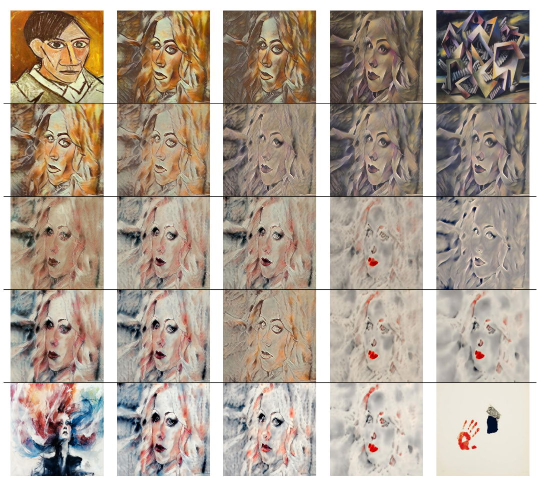

Table 3: Adaptive Instance Normalization Mixing Baseline

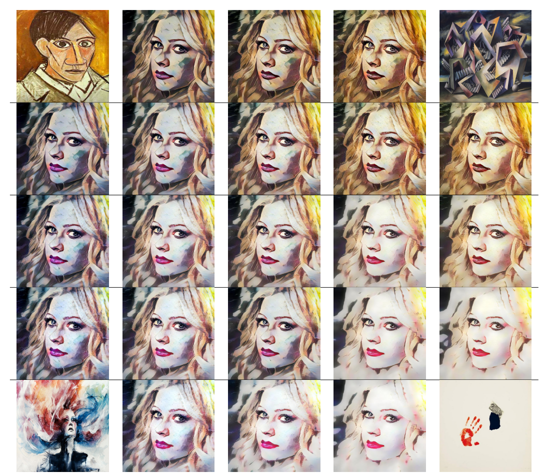

Figure 4: (Table 1): Wasserstein Barycenter Interpolation between a content image given above and four target styles images given at the corner of the square. Each image in the square is for an interpolation weight $(\lambda_1, \dots \lambda_4)$, that are defined on a grid on the square. (Table 2): Fisher Rao Interpolation between the same content image given above and the same four target styles images given at the corner of the square. In both cases Gaussian Wasserstein transport plans are used to obtain the transformed image to the novel mixed style in the feature space, and the final image is obtained using the decoder. (Table 3): the AdaIn baseline that we showed that it does a diagonal approximation fails at capturing the subtle details of the style of the target images. Both Wasserstein and Fisher Rao approaches are successful, we notice that while Wasserstein barycenter is color dominant in defining the new style, the Fisher Rao barycenter capture more the strokes and captures better color variations in the novel artistic styles. We note that the Wasserstein is smoother as we change the interpolation weights then the Fisher Rao. (Figure is better seen in color and zooming in; See Appendix for a full resolution).

target distribution in a feature space, can be cast as an optimal transport problem between Gaussian measures. The transport maps of Gaussian OT have closed forms and allow simple and efficient style transfer. Moreover, we showed how to mix different styles using Wasserstein barycenters between Gaussian measures. Mixing different styles using geodesic distances such as Wasserstein and the Fisher Rao metric allow better non linear interpolation, and gives rise to different stylizations of content images.

References

Agueh, M. and Carlier, G. Barycenters in the wasserstein space. *SIAM J. Math. Analysis*, 43, 2011.

Benamou, J.-D. and Brenier, Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 2000.

Bhatia, R. *The Riemannian Mean of Positive Matrices*. Springer Berlin Heidelberg, 2013.

Chen, T. Q. and Schmidt, M. Fast patch-based style transfer of arbitrary style, 2016.

Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pp. 2292–2300, 2013.

Dumoulin, V., Shlens, J., and Kudlur, M. A learned representation for artistic style. 2017. URL https://openreview.net/forum?id=BJO-BuT1g.

Ferradans, S., Papadakis, N., Rabin, J., Peyré, G., and Aujol, J.-F. Regularized discrete optimal transport. *Scale Space and Variational Methods in Computer Vision*, 2013.

Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S.-i., Trouvé, A., and Peyré, G. Interpolating between Optimal Transport and MMD using Sinkhorn Divergences. *ArXiv e-prints*, 2018.

Frogner, C., Zhang, C., Mobahi, H., Araya, M., and Poggio, T. A. Learning with a wasserstein loss. In *Advances in Neural Information Processing Systems*, pp. 2053–2061, 2015.

Gatys, L. A., Ecker, A. S., and Bethge, M. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016.

Gatys, L. A., Ecker, A. S., Bethge, M., Hertzmann, A., and Shechtman, E. Controlling perceptual factors in neural style transfer. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, 2017.

Huang, X. and Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.

Jing, Y., Yang, Y., Feng, Z., Ye, J., and Song, M. Neural style transfer: A review. *CoRR*, 2017.

Johnson, J., Alahi, A., and Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. *Lecture Notes in Computer Science*, 2016.

Kolkin, N., Salavon, J., and Shakhnarovich, G. Style transfer by relaxed optimal transport and self-similarity, 2019.

Li, C. and Wand, M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *ECCV (3)*, 2016a.

Li, C. and Wand, M. Combining markov random fields and convolutional neural networks for image synthesis. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016b. doi: 10.1109/cvpr.2016.272. URL http://dx.doi.org/10.1109/CVPR.2016.272.

Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., and Yang, M.-H. Universal style transfer via feature transforms. In *Advances in Neural Information Processing Systems 30*. 2017a.

Li, Y., Wang, N., Liu, J., and Hou, X. Demystifying neural style transfer. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017b.

Lu, M., Zhao, H., Yao, A., Chen, Y., Xu, F., and Zhang, L. A closed-form solution to universal style transfer. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

Marron, V. Style transfer as optimal transport. https://github.com/VinceMarron/style_transfer, 2018.

Masarotto, V., Panaretos, V. M., and Zemel, Y. Procrustes metrics on covariance operators and optimal transportation of gaussian processes. *Sankhya A*, 2018.

McCann, R. J. A convexity principle for interacting gases. *Advances in Mathematics*, 1997.

Peyré, G. and Cuturi, M. Computational optimal transport. Technical report, 2017.

Rabin, J., Peyré, G., Delon, J., and Bernot, M. Wasserstein barycenter and its application to texture mixing. In *Proceedings of the Third International Conference on Scale Space and Variational Methods in Computer Vision*, SSVM'11, 2012.

Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *CoRR arXiv:1409.1556*, 2014.

Takatsu, A. Wasserstein geometry of gaussian measures. *Osaka J. Math.*, 2011.

Ulyanov, D., Lebedev, V., Vedaldi, A., and Lempitsky, V. S. Texture networks: Feed-forward synthesis of textures and stylized images. In *ICML*, 2016.

Ulyanov, D., Vedaldi, A., and Lempitsky, V. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Wang, X., Oxholm, G., Zhang, D., and Wang, Y.-F. Multimodal transfer: A hierarchical deep convolutional neural network for fast artistic style transfer. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Wang, Z., Xia, G., and Zhang, Y. Texture mixing by interpolating deep statistics via gaussian models. *CoRR*, abs/1807.11035, 2018.

Wynen, D., Schmid, C., and Mairal, J. Unsupervised learning of artistic styles with archetypal style analysis. In *Advances in Neural Information Processing Systems*, pp. 6584–6593, 2018.

Xia, G., Ferradans, S., Peyré, G., and Aujol, J. Synthesizing and mixing stationary gaussian texture models. *SIAM J. Imaging Sciences*, 2014.

Zhang, H. and Sra, S. First-order methods for geodesically convex optimization. In *COLT*, 2016.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

Álvarez Esteban, P. C., del Barrio, E., Cuesta-Albertos, J., and Matrán, C. A fixed-point approach to barycenters in wasserstein space. *http://arxiv.org/pdf/1511.05355.*

Supplementary Material for Wasserstein Style Transfer

A Examples of Interpolating Content and Styles with Wasserstein Barycenter and Optimal Transport

In Figures 5, 6 ,7 we show examples of interpolations of content images with the style images. We used in this experiment a coarse to fine approach, i.e starting from matching upper layers of VGG to lower layers.

B Mixing Styles with Frechet Means and Optimal Transport Style Transfer

Coarse to Fine. -We give results of different Mixing strategies and a content stylization in a coarse to fine procedure as follows: Wasserstein Mixing in Table 1;Fisher Rao Mixing in Table 2 ;Arithmetic Mixing that would be close to WCT baseline (Li et al., 2017a) in Table 3; Harmonic Mixing in Table 4 ; AdaIN Mixing Table in 5. We also give another set of results on Wass Barycenter mixing in Table 8, Fisher Rao in Table 9 and AdaIn in 10

Fine to Coarse. We experiment baselining WCT mixing (Li et al., 2017a) and Wasserstein Mixing in a Fine to coarse strategy (from lower layer to upper layers) results are given in Table 6 and Table 7.

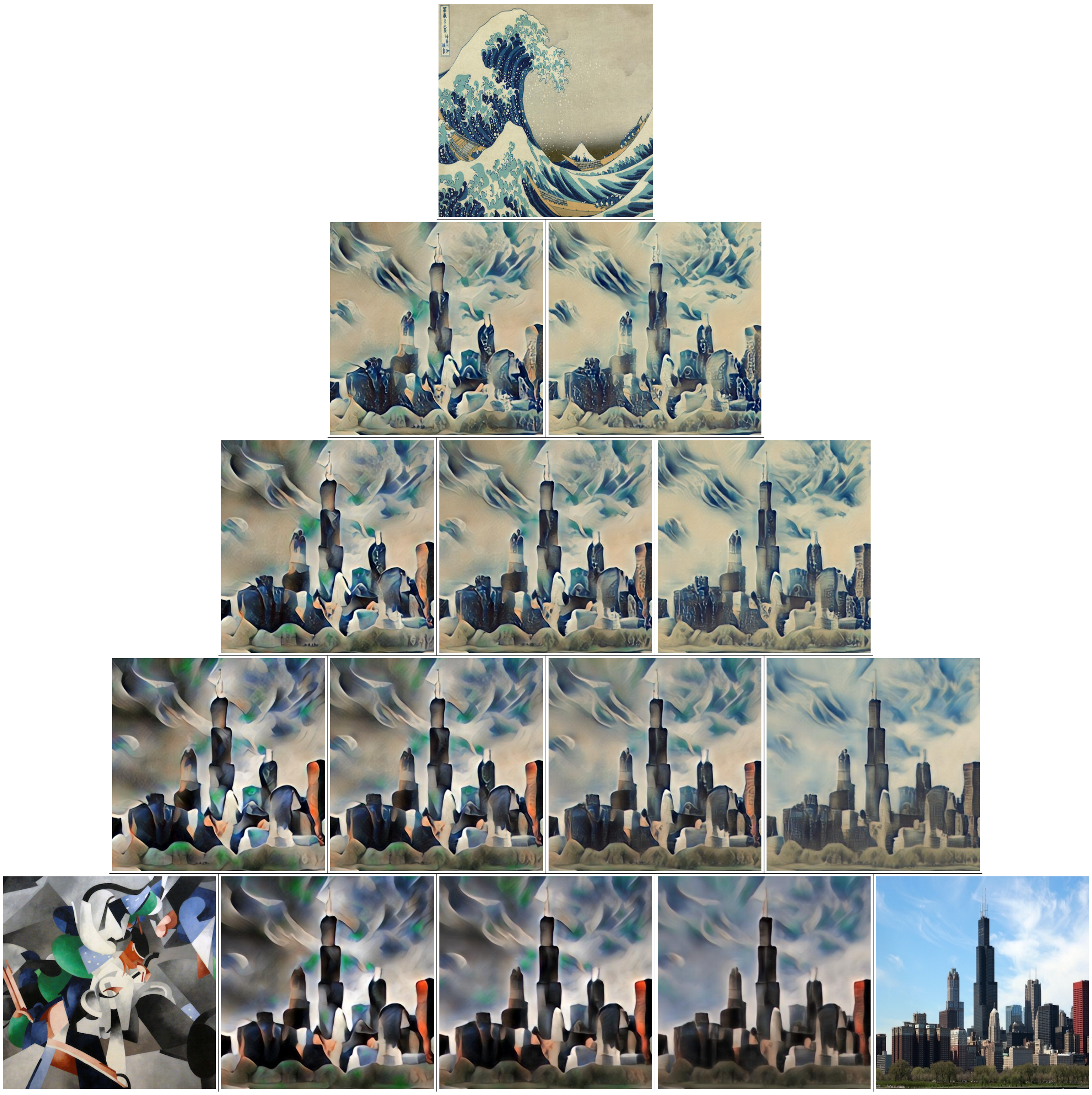

Figure 5: Wasserstein barycenters for Style Mixing and Transfer. The content image on the right corner of the triangle is mixed with the two styles images. Each image in the triangle correspond to a set of interpolation weights defined by proximity to the content or style images in the triangle.

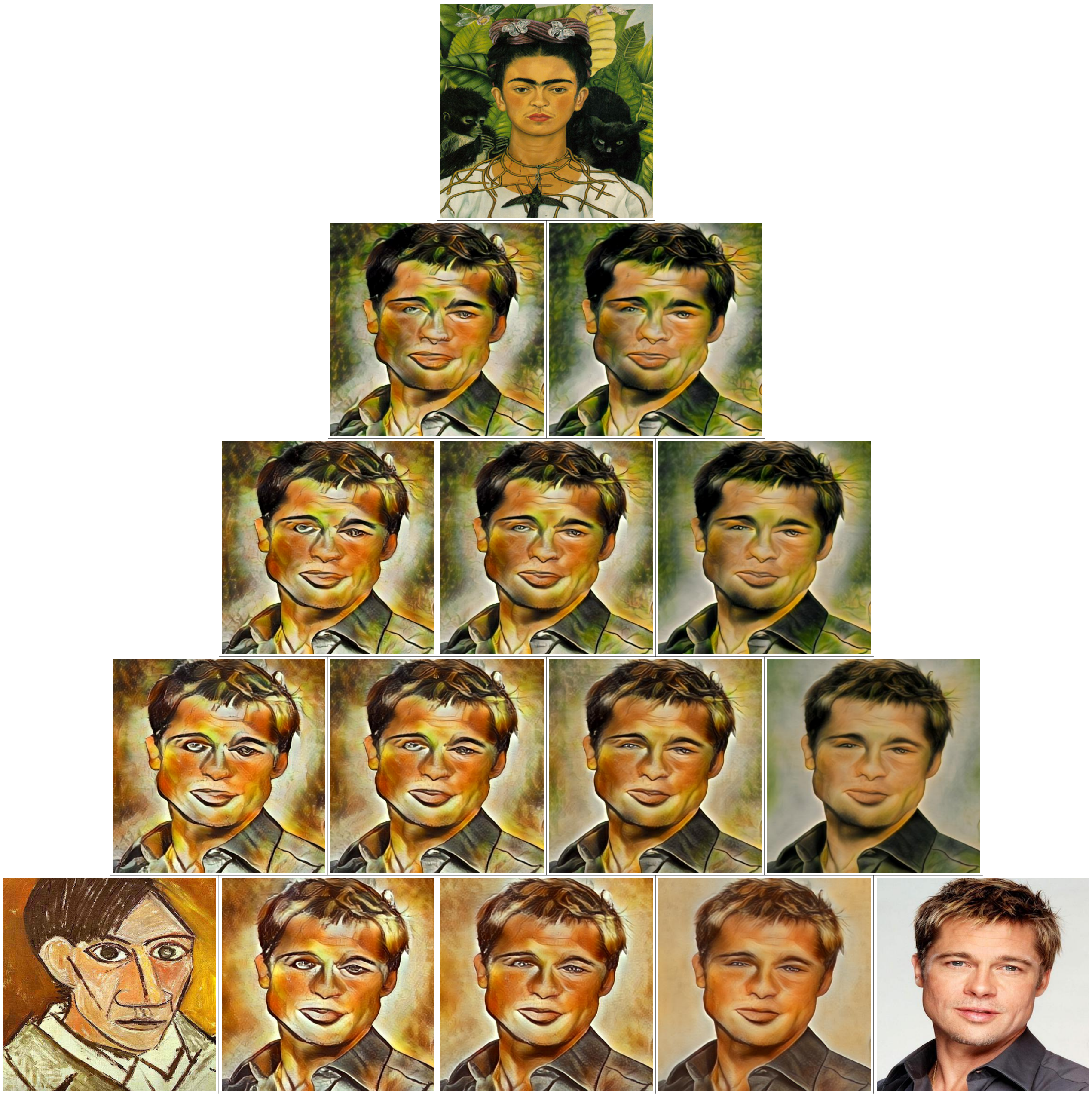

Figure 6: Wasserstein barycenters for Style Mixing and Transfer. The content image on the right corner of the triangle is mixed with the two styles images. Each image in the triangle correspond to a set of interpolation weights defined by proximity to the content or style images in the triangle.

Figure 7: Wasserstein barycenters for Style Mixing and Transfer. The content image on the right corner of the triangle is mixed with the two styles images. Each image in the triangle correspond to a set of interpolation weights defined by proximity to the content or style images in the triangle.

Figure 8: Content Image. We give results of different Mixing strategies and a content stylization in a coarse to fine procedure as follows: Wasserstein Mixing in Table 1;Fisher Rao Mixing in Table 2 ;Arithmetic Mixing that would be close to WCT baseline (Li et al., 2017a) in Table 3; Harmonic Mixing in Table 4 ; AdaIN Mixing Table in 5; The style images are given on the four corners of each square.

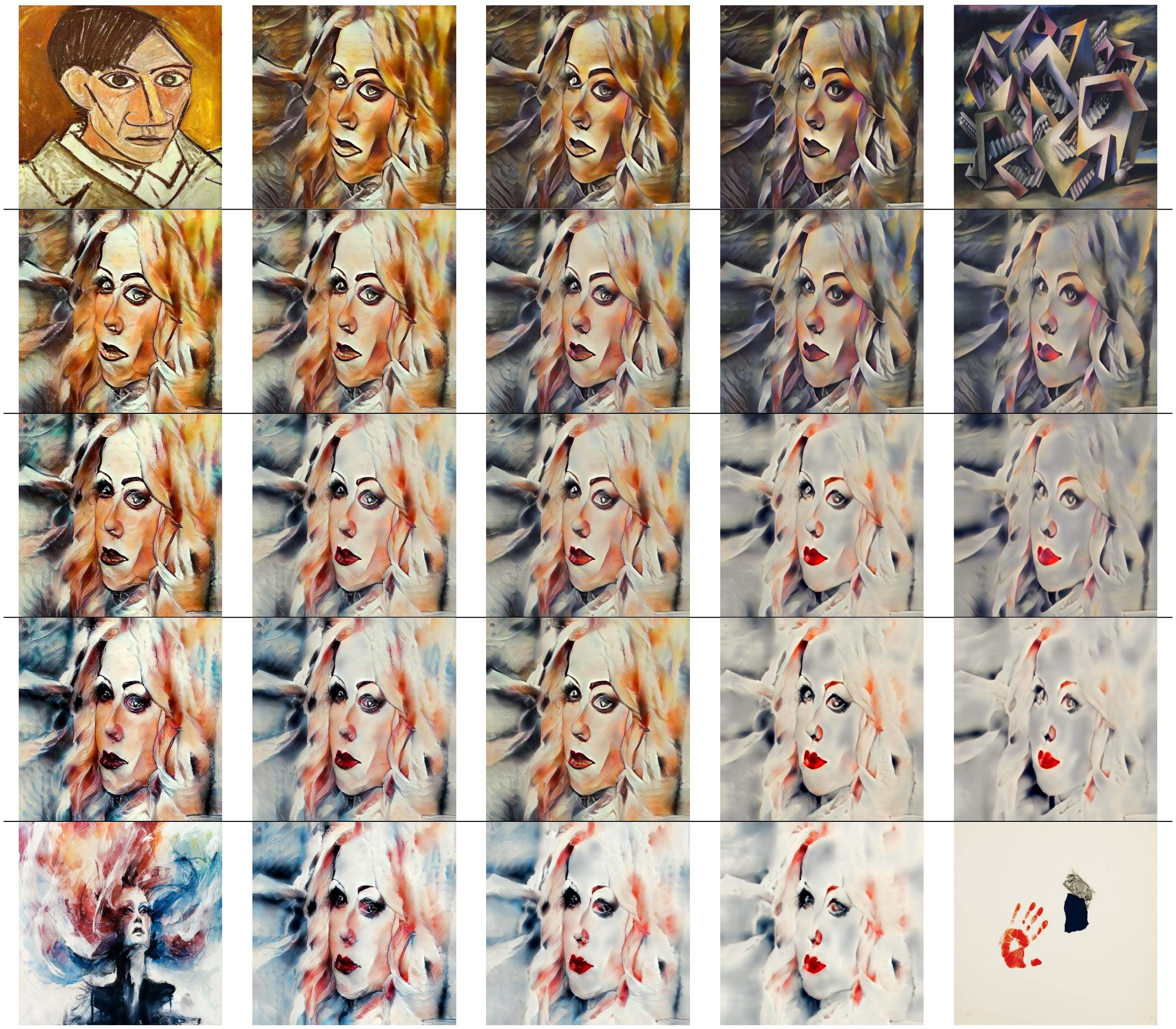

Table 1: Coarse to Fine style Transfer: Content image is given in Figure 8. Wasserstein Barycenter Mixing of the styles (the four images in the corners of the square) and Wasserstein Transport of the content image to the novel style defined by the Wasserstein Barycenter for various interpolations weights. The stylized image is generated by following a coarse to fine scheme.

Table 2: Coarse to Fine Generation: Karcher (Fisher Rao) Barycenter Mixing of the styles and Wasserstein Transport of the content image to the novel style defined by the Fisher Rao barycenter. The stylized is generated following a coarse to fine scheme.

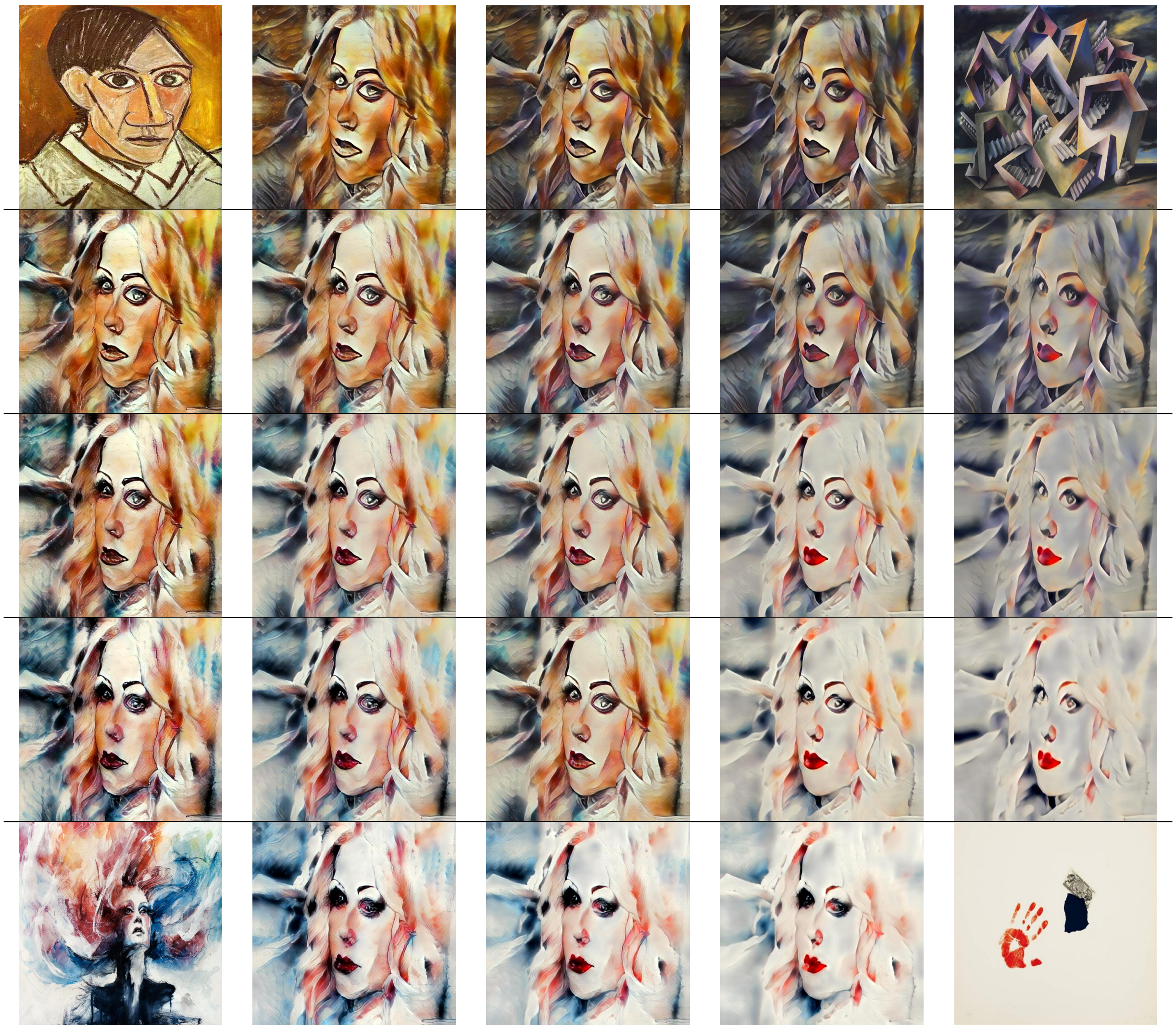

Table 3: Coarse Fine Style Transfer: (Arithmetic)Euclidean Barycenter Mixing of Covariances and Wasserstein Transport. This is similar to a WCT type of mixing. The stylized image is similar to a Wasserstein Barycenter Mixing, nevertheless a closer look shows subtle differences. This hints to the fact the coarse layers are almost diagonal.

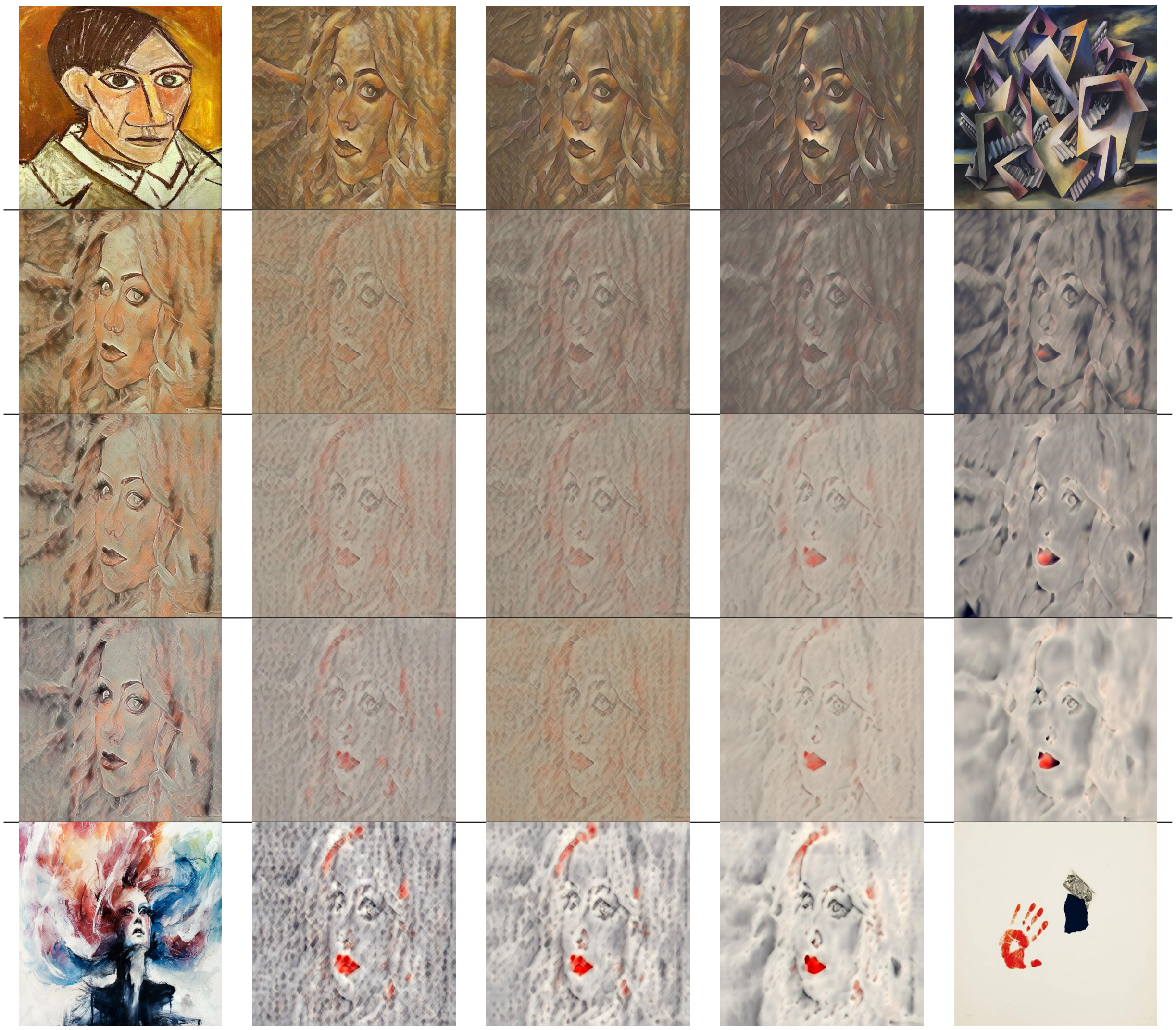

Table 4: Coarse Fine Style Transfer: Harmonic Barycenter Mixing of Covariances and Wasserstein Transport. The Harmonic Mixing have saturation problems and does not produce good results.

Table 5: Adaptive Instance Normalization Mixing Baseline (Huang & Belongie, 2017).

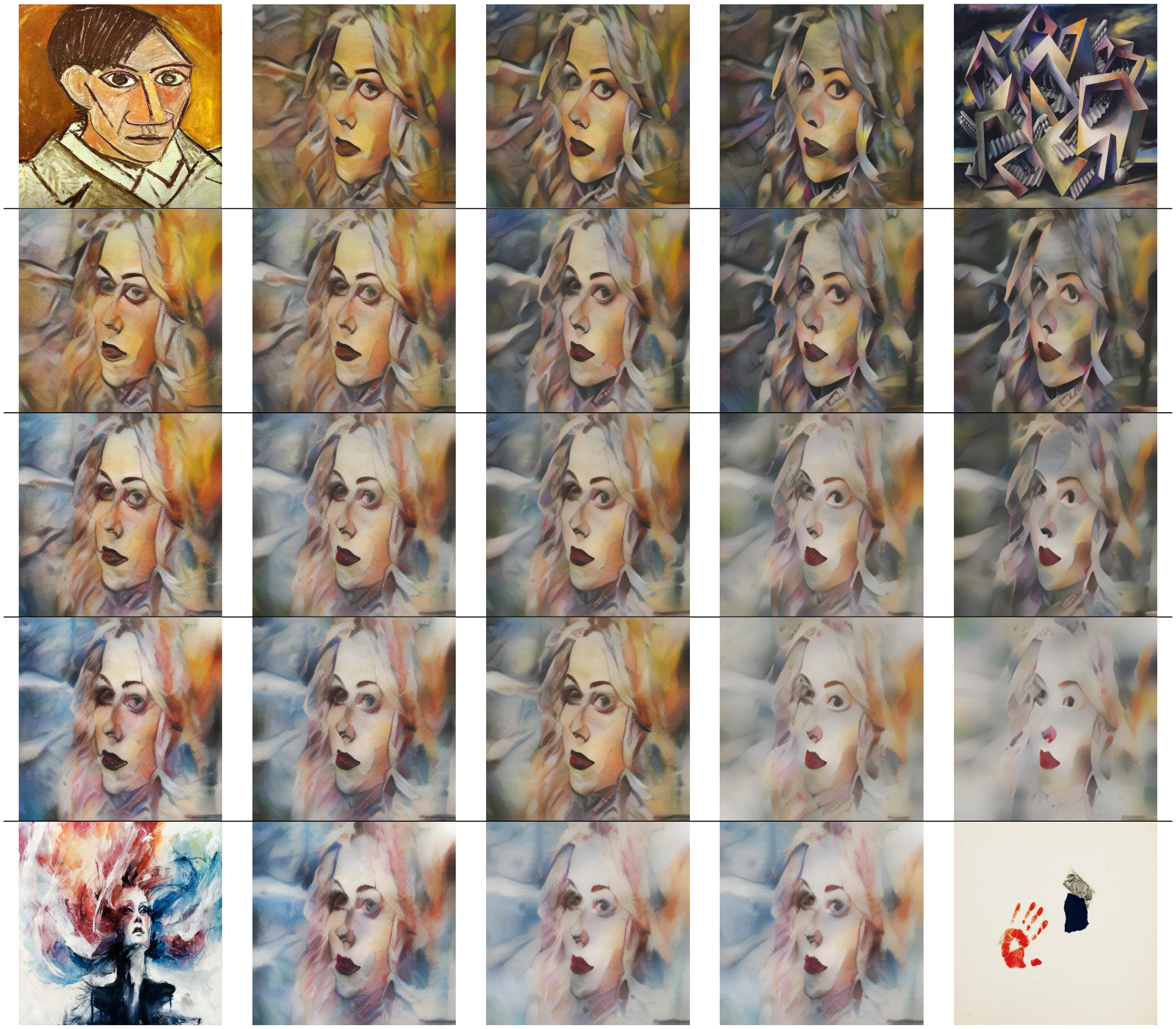

Table 6: Fine to coarse style Transfer: Content image is given in Figure 8. Wasserstein Barycenter Mixing of the styles (the four images in the corners of the square) and Wasserstein Transport of the content image to the novel style defined by the Wasserstein Barycenter for various interpolations weights. The stylized image is generated by following a coarse to fine scheme.

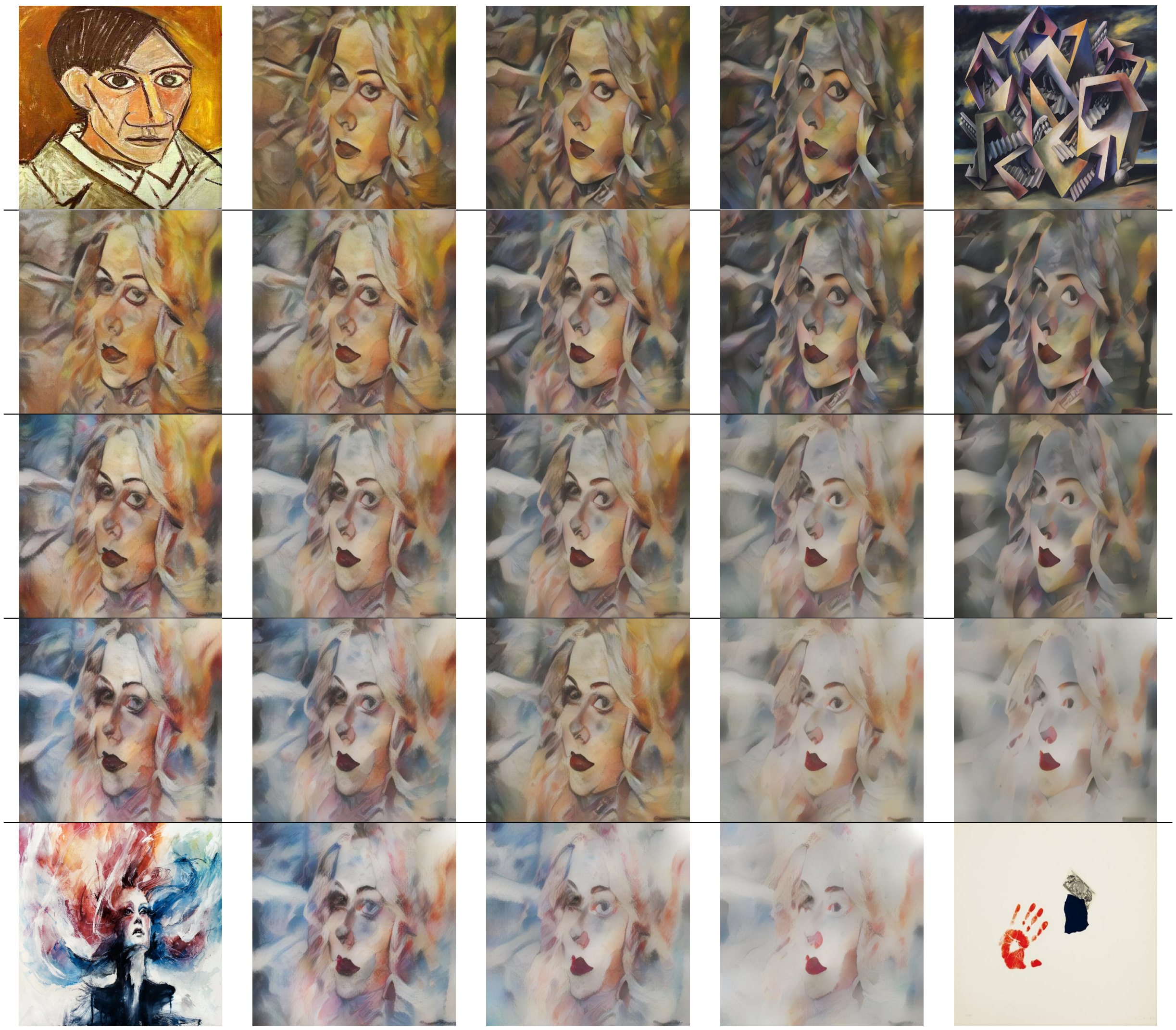

Table 7: Fine to coarse style Transfer: Results of Arithmetic Mean Mixing (WCT Mixing (Li et al., 2017a)) , we see here that the coloring has a lot of black shadow over the face unlike the Wasserstein barycenter mixing approach, in the previous Table.

Figure 9: Content Image (a Photo of Hotel Dieu painted by van gogh in the most right corner of the square). Stylization in mixture of four styles incluing van gogh painting are in Tables 8 for Wasserstein mixing and Table 9 for Fisher Rao Mixing. Table 10 is the AdaIN baseline.

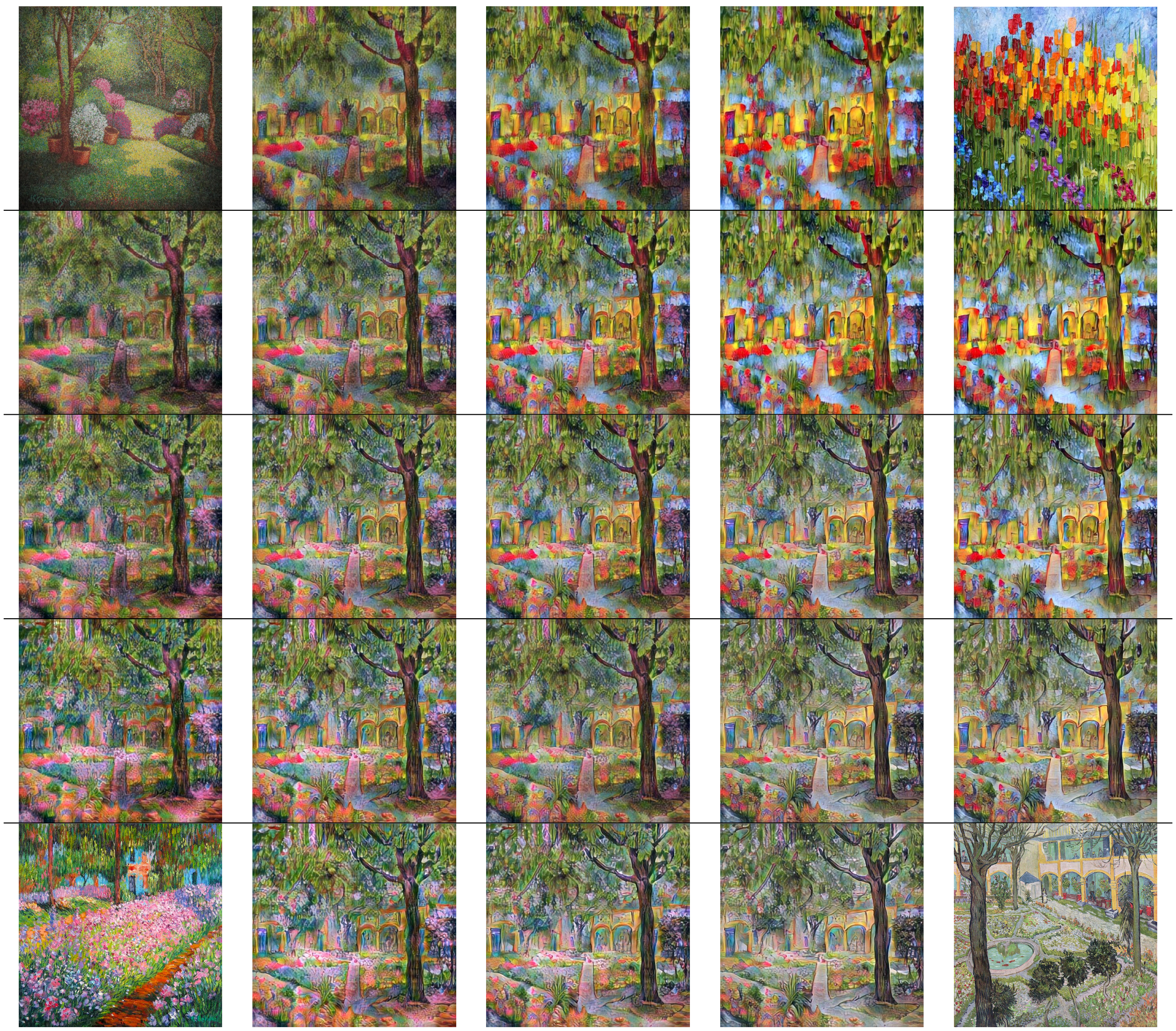

Table 8: Coarse to Fine style transfer: Wasserstein Barycenter Mixing and Wasserstein Transport. The content image is given in Figure 9. The four styles are on the four corner of the square.

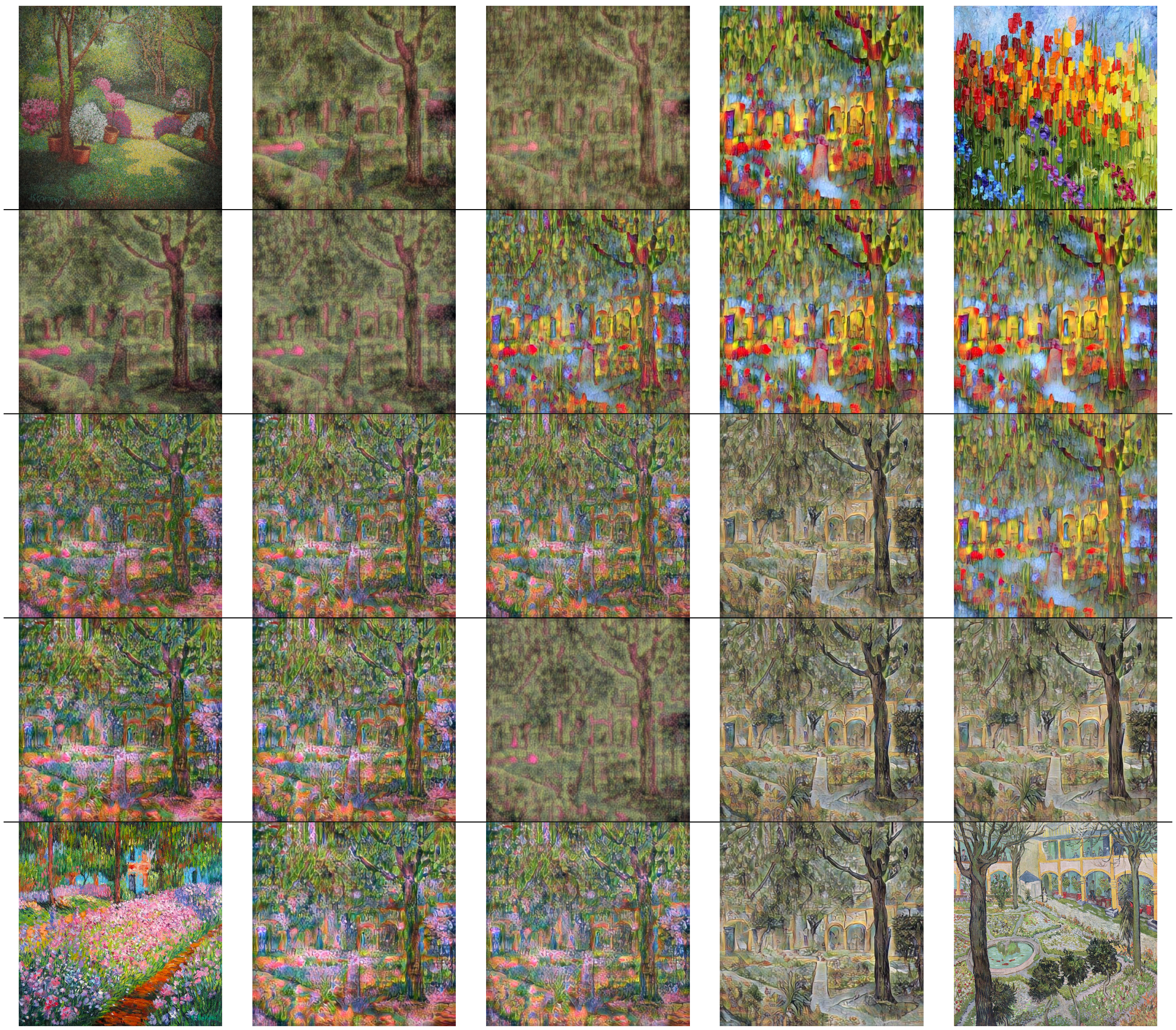

Table 9: Coarse to fine style transfer: Fisher Rao Mixing and Wasserstein Transport.

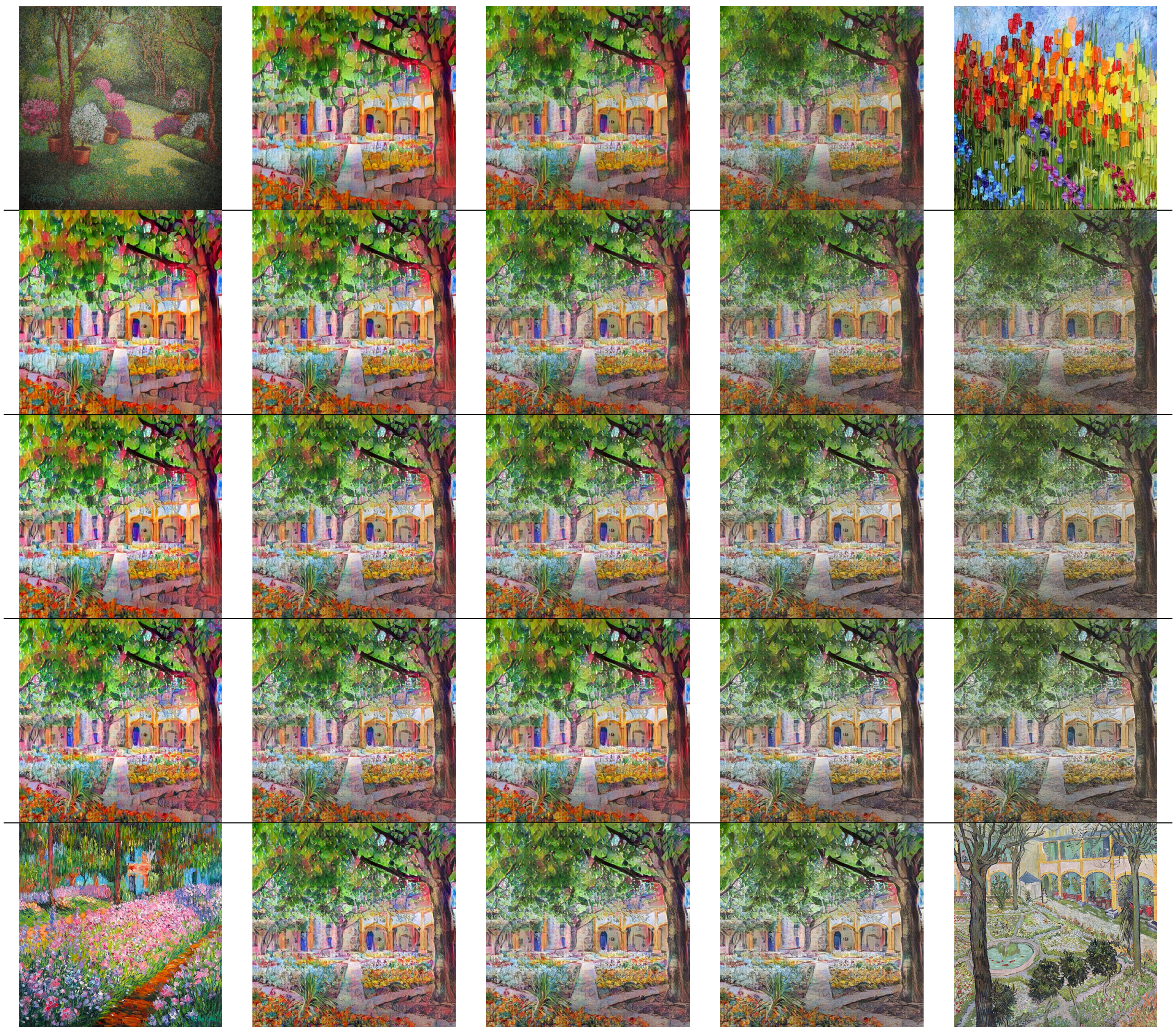

Table 10: Adaptive Instance Normalization Mixing Baseline (Huang & Belongie, 2017).