IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Style Transfer by Relaxed Optimal Transport and Self-Similarity

Group 5:
Houda GHALLAB
Renzo MORALES
Carlos ARGUILAR
Mohamed Salim ARIFA
Ghaith MAKHLOUF

SUMMARY

1. Introduction
2. Fundamentals : Optimal Transport & Self Similarity
3. Methodology
4. Implementation and region control
5. Experiments and comparison to the related work
6. Conclusion

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

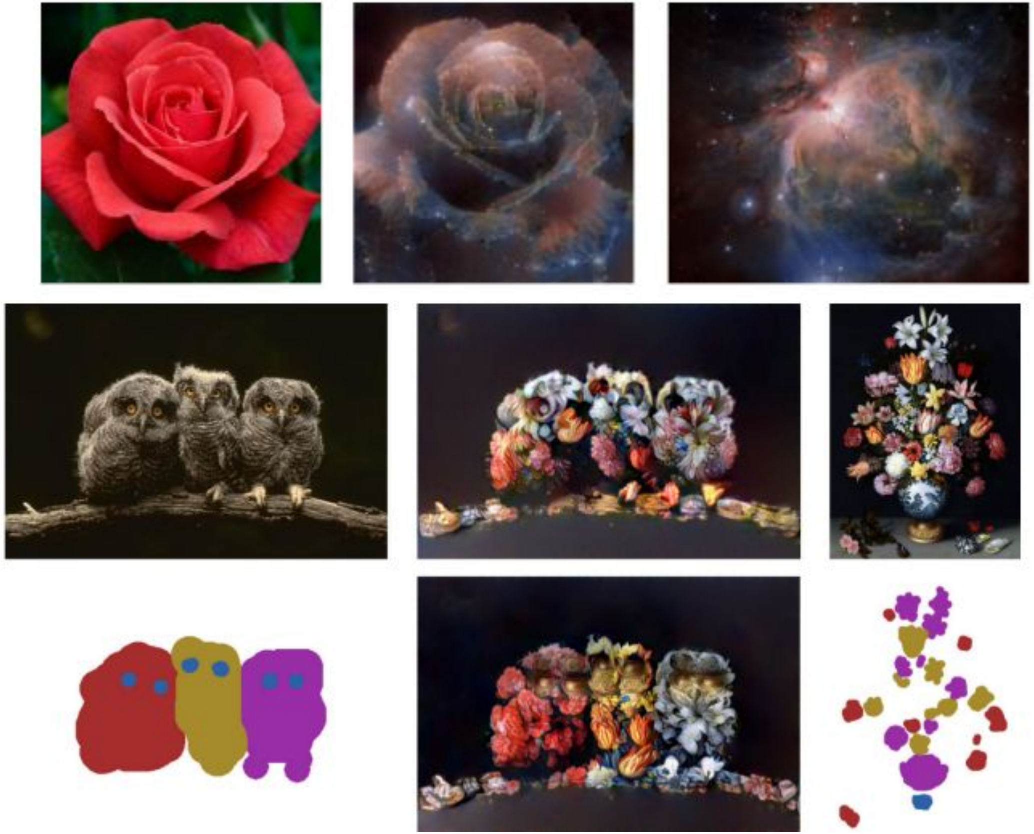
INTRODUCTION

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

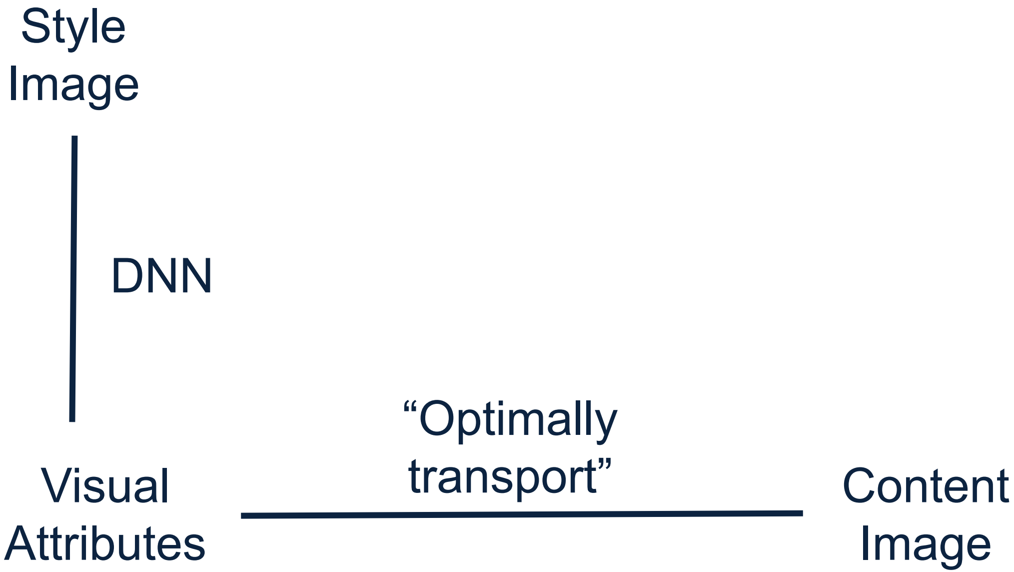
Content

Style

Combination



Style: a distribution over features extracted by a deep neural network



Style: a distribution over features extracted by a deep neural network

Style Image

DNN

Visual Attributes

"Optimally transport"

Content Image

with minimum distortion to the content

Content:

Self similarity: objects often have repeating patterns or elements within themselves

Content:

Human visual system: Relative appearance and surroundings

Content:

Self similarity: objects often have repeating patterns or elements within themselves

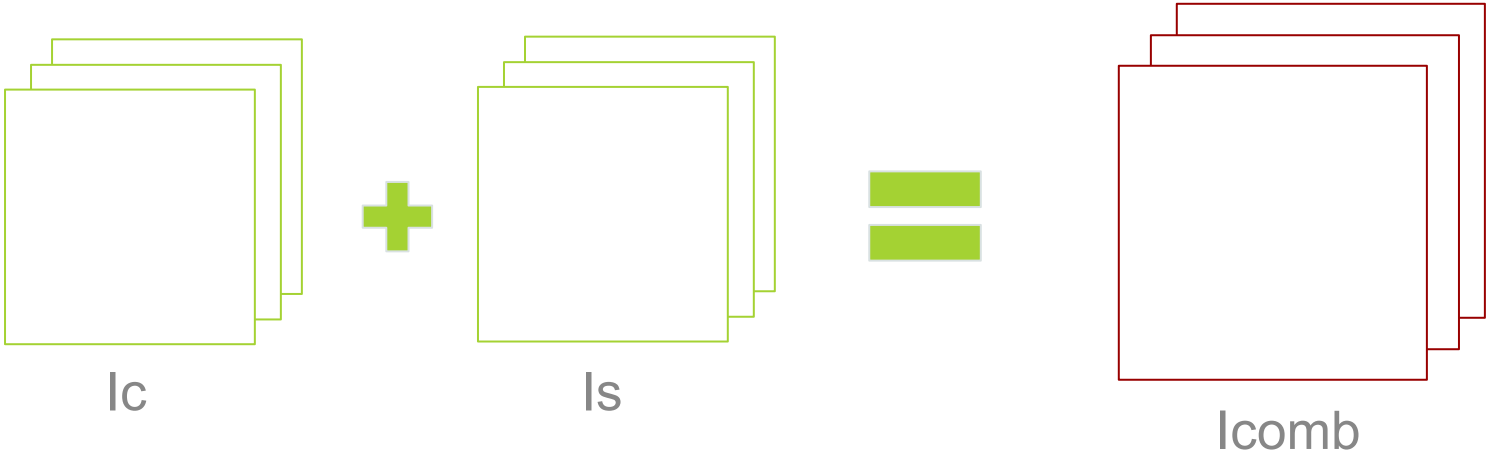
Human visual system: Relative appearance and surroundings

→

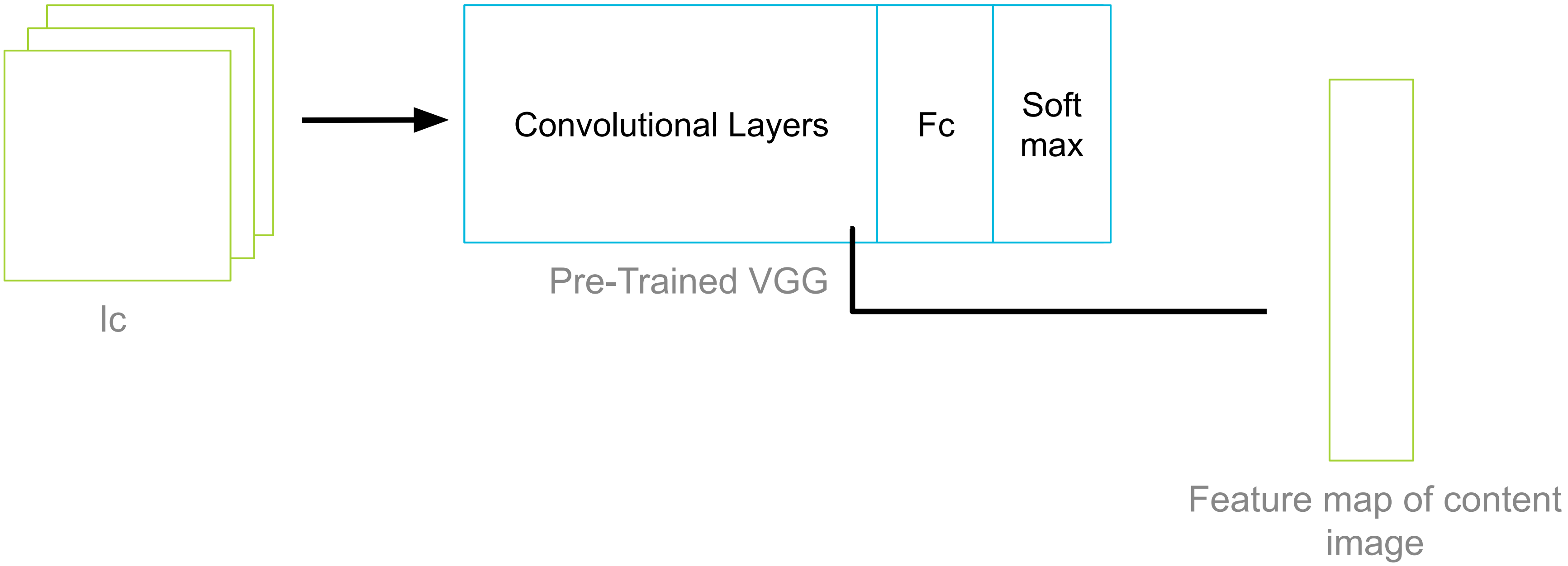
Preserve **semantics** and **spatial** layout of the content image

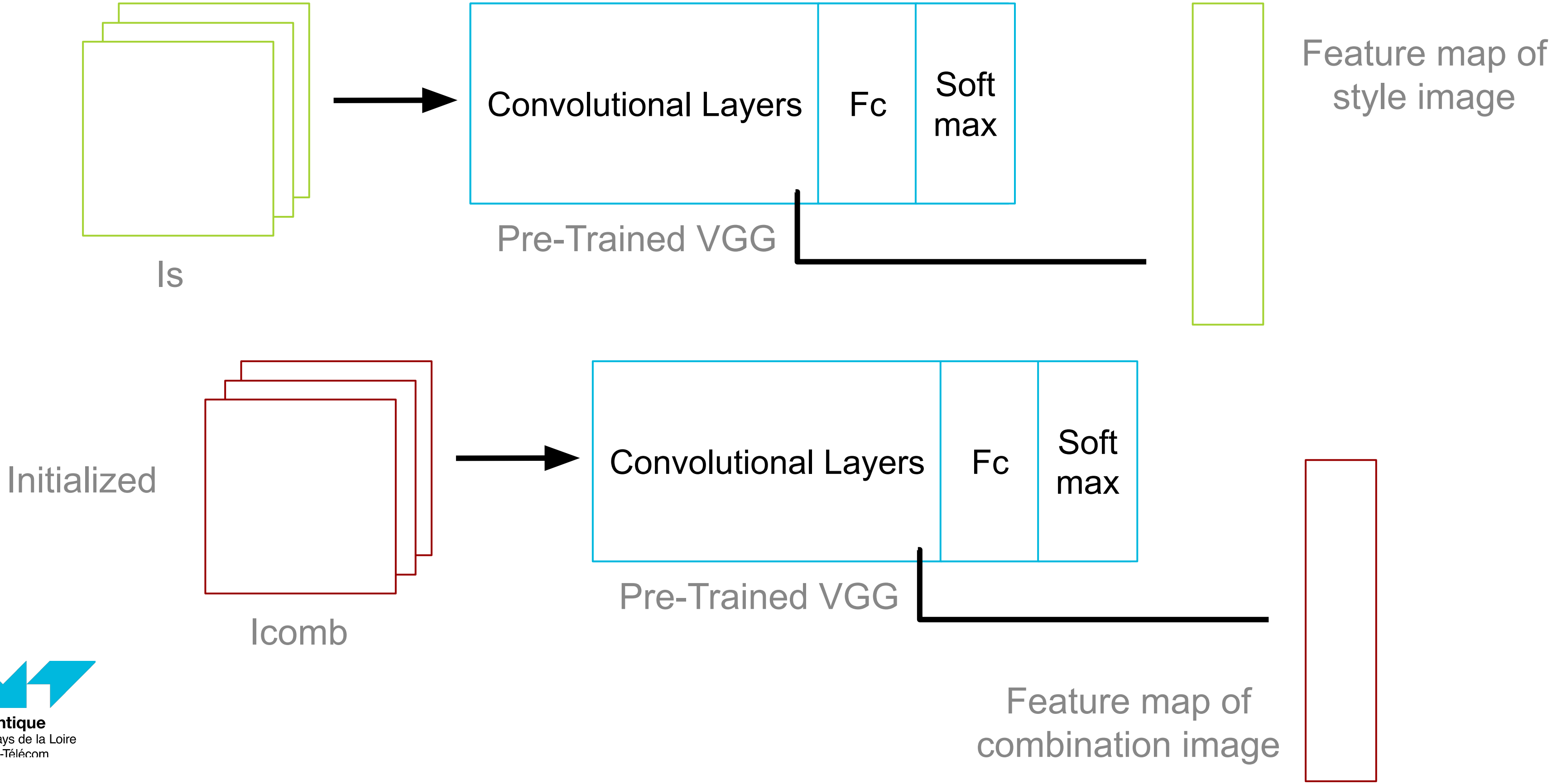
Content is not focused on the exact pixel values of the image, but rather on the relationships between those pixels.

1.3 Algorithm overview



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom


Convolutional Layers
Fc
Soft
max
Pre-Trained VGG
Ic
Feature map of content
image


Convolutional Layers
Fc
Soft
max
Feature map of
style image
Pre-Trained VGG
Is
Initialized
Convolutional Layers
Fc
Soft
max
Pre-Trained VGG
Icomb
Feature map of
combination image

Features extraction: **feature maps** of **content**, **style**, and **combination** images

Optimization: We want to measure how far our combination image is from a “Good combination”

Minimization of a loss function
Icomb_Loss = Loss(Ic) + Loss(Is)

Features extraction: **feature maps** of **content**, **style**, and **combination** images

Gradient descent

```
init Icomb
for e in range(epochs):
    calculate Icomb_loss
    derivative of loss
    update Icomb accordingly
```

FUNDAMENTALS

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

- Let's break down the title of the paper:

Style Transfer by Relaxed Optimal Transport and Self-Similarity

- Let's break down the title of the paper:

Style Transfer by Relaxed Optimal Transport and Self-Similarity

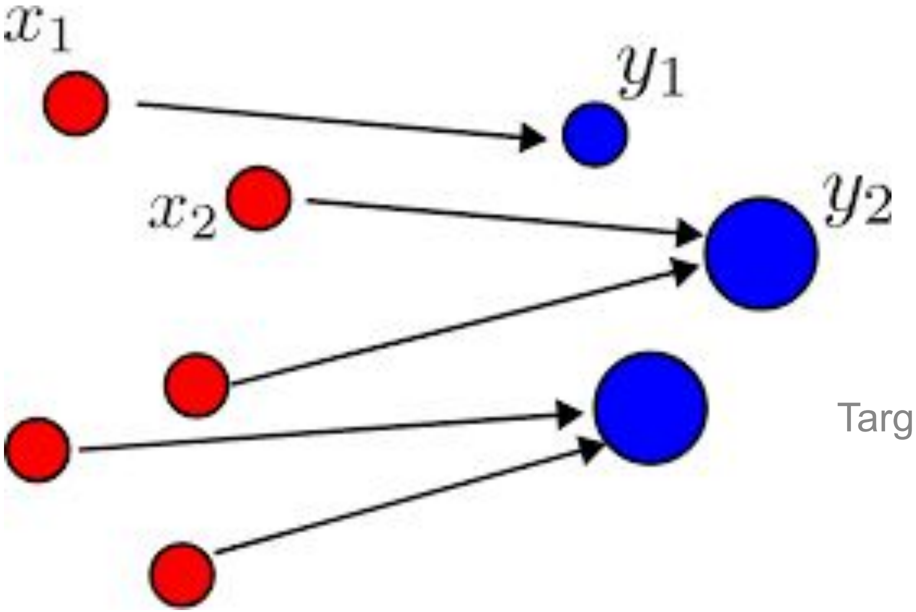
Optimal transport:

Measures: initial distribution of "mass" in a space and where we want to move it to

Cost function: cost of moving mass between points in the source and target spaces

Transport map: how to move the mass from the source distribution to the target distribution while minimizing the total cost

Source Space



Target Space

Optimal transport:

Given two probability measures u (source) and v (target) on a Polish space (a complete, separable metric space) X, and a cost function c: X x X → ℝ, the optimal transport problem seeks a transport map T: X → X that minimizes the total cost of transporting the mass:

don't know if
is slide is
ecessary

∫_X c(x, T(x)) dμ(x)
subject to the constraints that T pushes forward μ to v (i.e., for any measura
set A in X, v(A) = μ(T^(-1)(A)))

Solution: Earth mover's distance EMD

$$C_{ij} = D_{\cos}(A_i, B_j) = 1 - \frac{A_i \cdot B_j}{\|A_i\| \|B_j\|}$$

$$\mathrm{EMD}(A, B) = \min_{T \geq 0} \sum_{ij} T_{ij} C_{ij} \qquad (2)$$

$$s.t. \sum_j T_{ij} = 1/m \qquad (3)$$

$$\sum_i T_{ij} = 1/n \qquad (4)$$

A and B: Sets of vectors
T: Transport matrix
C: Cost matrix
n: len(A)
m: len(B)

$$\boxed{O(\max(m, n)^3)}$$

Exact match between features

Approximation: Relaxed earth mover's distance REMD

$$R_A(A,B) = \min_{T \geq 0} \sum_{ij} T_{ij} C_{ij} \quad s.t. \quad \sum_j T_{ij} = 1/m \quad (5)$$

$$R_B(A,B) = \min_{T \geq 0} \sum_{ij} T_{ij} C_{ij} \quad s.t. \quad \sum_i T_{ij} = 1/n \quad (6)$$

A and B: Sets of vectors
T: Transport matrix
C: Cost matrix
n: len(A)
m: len(B)

Approximation: Relaxed earth mover's distance REMD

$$\ell_r = REMD(A, B) = \max(R_A(A, B), R_B(A, B)) \quad (7)$$

This is equivalent to:

$$\ell_r = \max\left(\frac{1}{n}\sum_i \min_j C_{ij}, \frac{1}{m}\sum_j \min_i C_{ij}\right) \quad (8)$$

A and B: Sets of vectors
T: Transport matrix
C: Cost matrix
n: len(A)
m: len(B)

→ More imperfections in feature matching, more natural looking style transfer

Ideas to explain

- what is self similarity
- the mathematical basis and proofs used
- Link between theory and
- In id tortor sit amet augue.

#here the need for a self similarity descriptor to preserve the original content is described

Trying to define Self-Similarity

Content Image

Style image

Undesirable output and total loss of content

Trying to define Self-Similarity

= Same "content"
Same internal self similarity

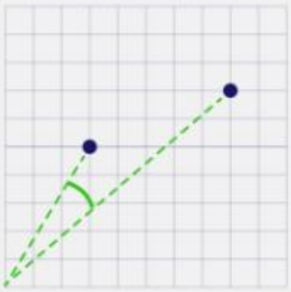
≠ different colours
different edges
different texture

another example of content preservation: Pareidolia

#trying to define what "content" is, impiortantly to say that the structure cannot be perceived logically but should be analysed within the context of an image

Solution:

Self similarity descriptor=

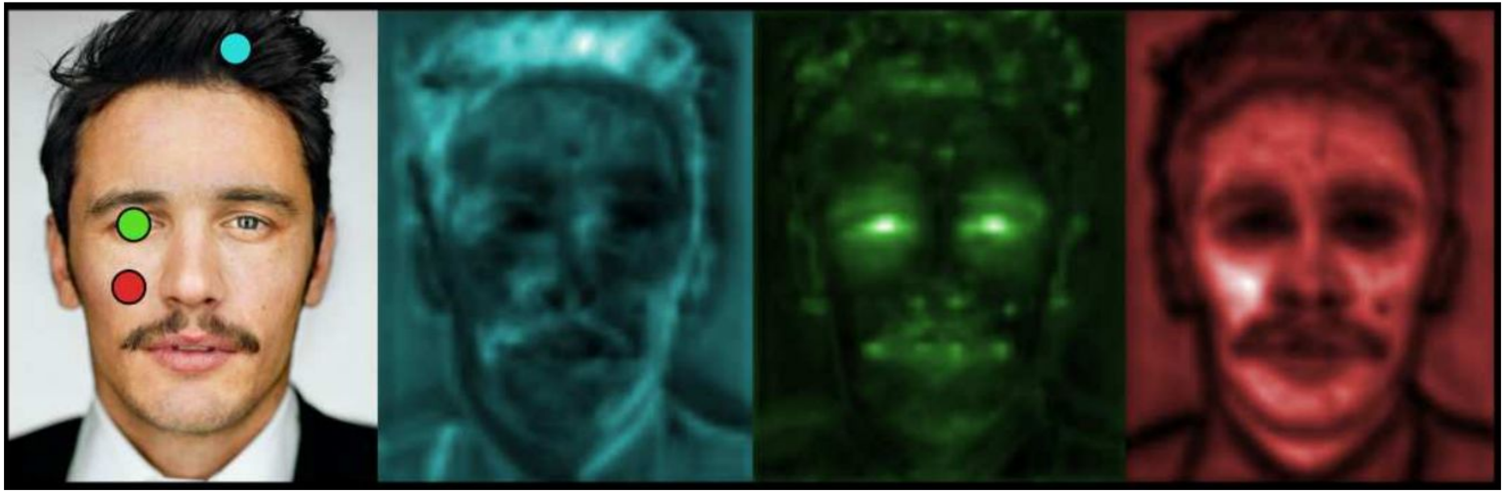


Cosine Distance

$$1 - \frac{A \cdot B}{||A|| \ \ ||B||}$$

doesn’t care about:

- -luminance
- -texture
- -scale
- -Colour



Self similarity Map (1- cosine_distance)

$$D_{ij} = \begin{array}{cc} & \begin{array}{cccc} a & & b & c \end{array} \\ \begin{array}{c} a \\ \\ b \\ c \end{array} & \begin{bmatrix} 0 & \cdots & 0.3 & 0.84 \\ \vdots & 0 & \vdots & \vdots \\ 0.3 & \cdots & 0 & \vdots \\ 0.84 & \cdots & \cdots & 0 \end{bmatrix} \end{array}$$

Self similarity Matrix (of cosine distances)

METHODOLOGY

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Loss Function

How are optimal transport and style similarity concepts used in our work ?

Loss Function :

$$L(X, I_C, I_S) = \frac{\alpha \ell_C + \ell_m + \ell_r + \frac{1}{\alpha}\ell_p}{2 + \alpha + \frac{1}{\alpha}}$$

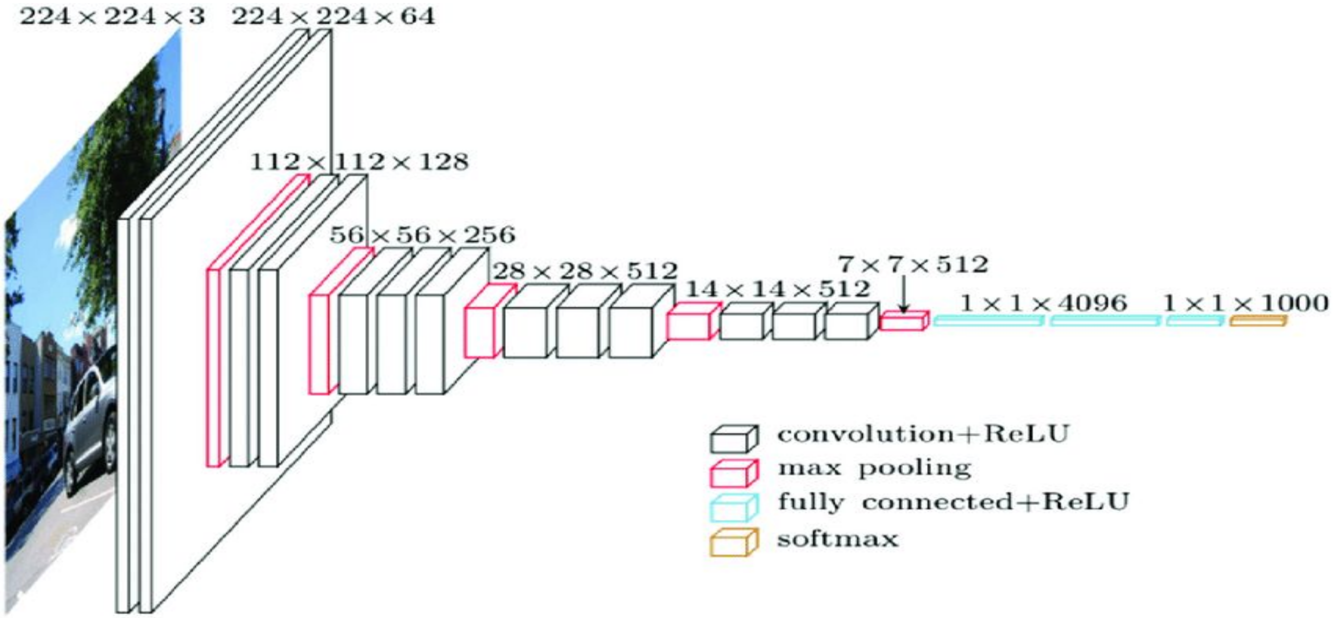
Loss Function

$$L(X, I_C, I_S) = \frac{\underbrace{\alpha \ell_C}_{\text{Content Loss}} + \underbrace{\ell_m + \ell_r + \frac{1}{\alpha}\ell_p}_{\text{Style Loss}}}{2 + \alpha + \frac{1}{\alpha}}$$

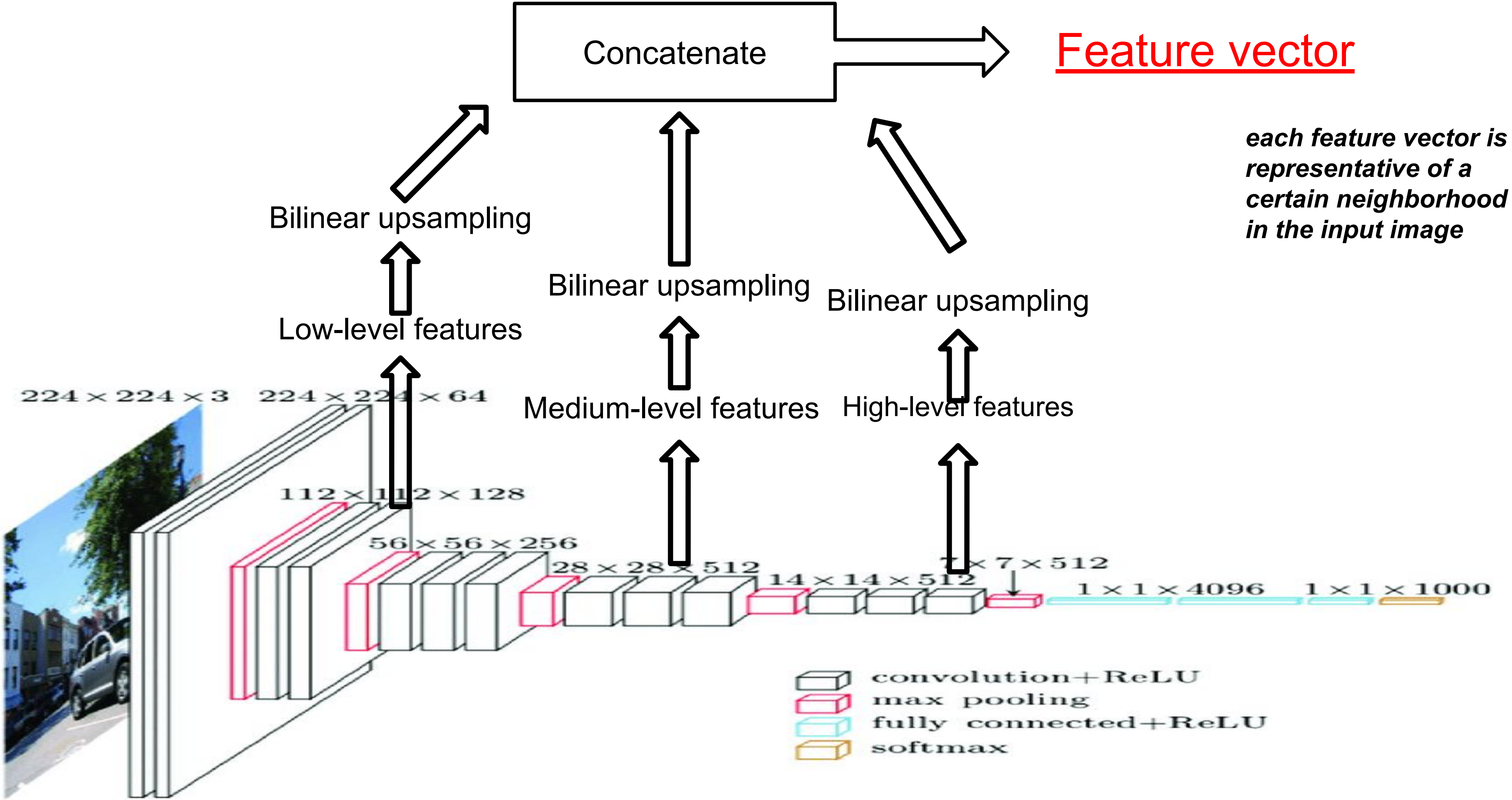
α: Hyperparameter: relative importance of content preservation to stylization

Both our style and content loss terms rely upon extracting a rich feature representation from an arbitrary spatial location.

VGG16 trained on ImageNet



VGG-16 for feature extraction



each feature vector is representative of a certain neighborhood in the input image

Content Loss

$$\mathcal{L}_{content}(X, C) = \frac{1}{n^2}\sum_{i,j}\left|\frac{D^X_{ij}}{\sum_i D^X_{ij}} - \frac{D^{I_C}_{ij}}{\sum_i D^{I_C}_{ij}}\right|$$

- D^X : The pairwise cosine distance matrix of all feature vectors extracted from X(t).
- D^{Ic} : The pairwise cosine distance matrix of all feature vectors extracted from Ic

Normalized cosine distance

Self-Similarity

$$\ell_m + \ell_r + \frac{1}{\alpha}\ell_p$$

Magnitude

Optimal Transport

Palette (colors)

Optimal Transport Problem

EMD ? $\Longrightarrow$ Costly: O(max(m,n)^3) $\Longrightarrow$ R-EMD

$$\ell_r = \max\left(\frac{1}{n}\sum_i \min_j C_{ij}, \frac{1}{m}\sum_j \min_i C_{ij}\right)$$

C: Cost Matrix: How far an element of A (set of feature vectors extracted from X) is from an element of B (set of feature vectors extracted from Is)

ℓ_r : Good transfer of the structural forms of the source image to the target. BUT..

ℓ_p : Color matching loss (palette) $\Longrightarrow$ R-EMD between X and Is

Magnitude of the feature vectors ignored by cosine distance $\Longrightarrow$ Visual Artifacts (under/over saturation)

Palette shifting is at odds with content preservation

$\Downarrow$

$$\frac{1}{\alpha}\ell_p$$

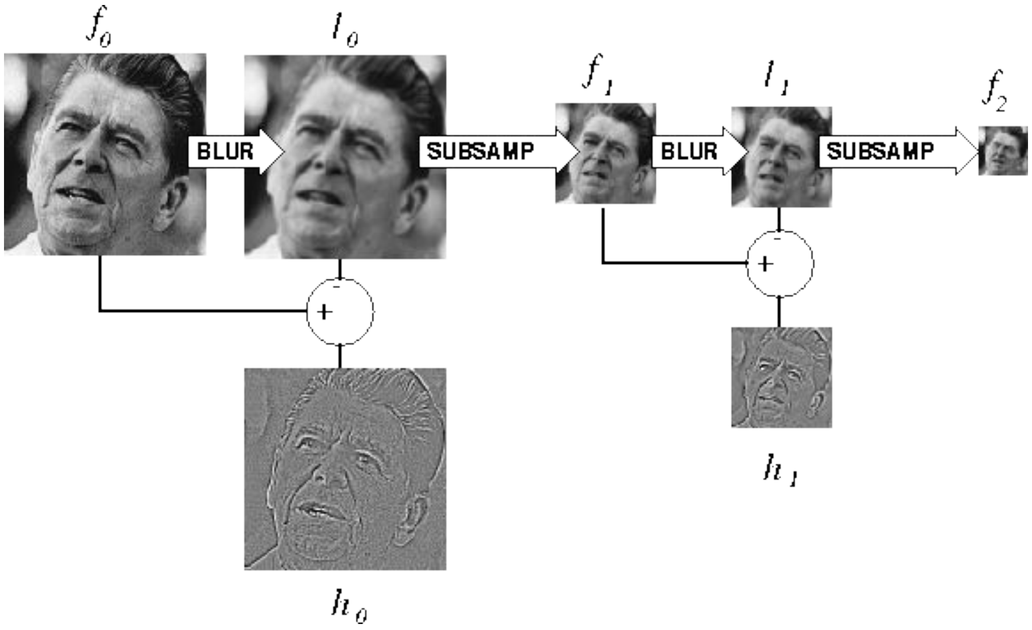
$$\ell_m = \frac{1}{d}\|\mu_A - \mu_B\|_1 + \frac{1}{d^2}\|\Sigma_A - \Sigma_B\|_1$$

- μ_A (μ_B) and **Σ**_A (**Σ**_B) are the mean and covariance of the feature vectors in set A (in set B).

IMPLEMENTATION AND REGION CONTROL

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Laplacian Pyramid: Image representation consisting of band-pass images and low-frequency residual image:



Goal:

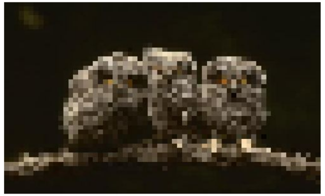
Content Image

Style Image

Stylized Image

STEP 0: INITIALIZATION

Δ (



Content Image (downscaled)

) + MC (

Style Image

) =

Stylized Image

Δ = **Laplacian**

MC = **Mean Color**

IMPLEMENTATION AND USER CONTROL

Step by Step

STEP 1: LAPLACIAN PYRAMID DE COMPOSITION

STEP 2: LOSS CALCULATION

L (


Content Image (downscaled)

,


Style Image (downscaled)

,

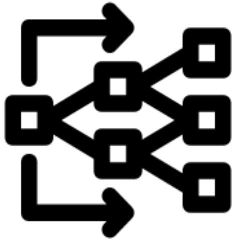

Stylized Image

)

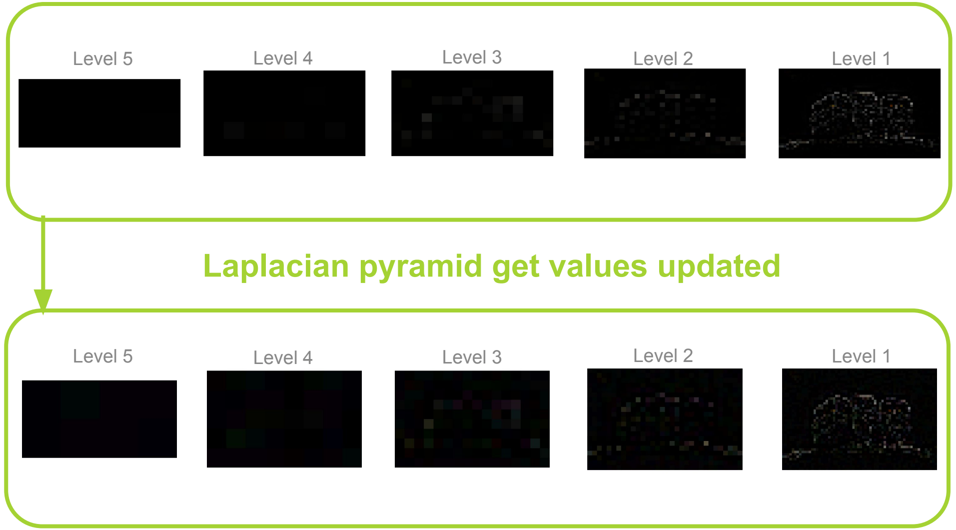
L = **Loss function**

STEP 2: BACKPROPAGATE AND UPDATE PYRAMID VALUES

$$L\left(\ \ldots\ \right)$$



Backpropagation



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

L **= Loss function**

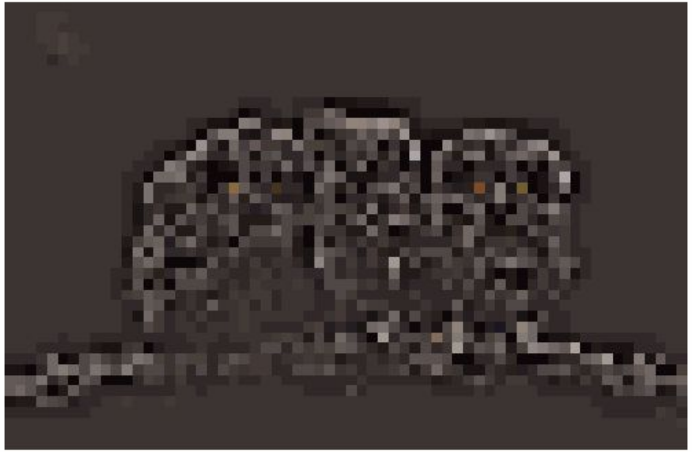
STEP 3: RECONSTRUCT IMAGE FROM UPDATED PYRAMID

Laplacian pyramid with updated values

(new) Stylized Image

Repeat step 1 to 3 (250 iterations) using the updated stylized image

Scale: 64 px, Iteration: 0



IMPLEMENTATION AND USER CONTROL

Multiple scale calculation

The iterative process described previously is made for 4 different scales using the (upscaled) output of the previous scale as input, halving the content weight (α) for the next scale:

long side 64px

α

long side 128px

$\alpha/2$

long side 256px

$\alpha/4$

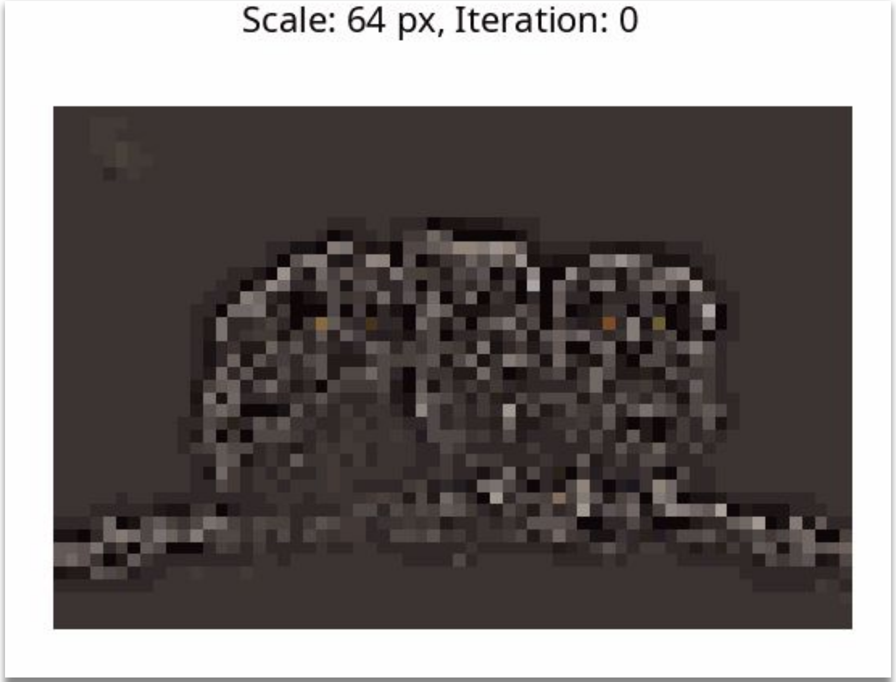
long side 512px

$\alpha/16$

Optimization followed by upscaling

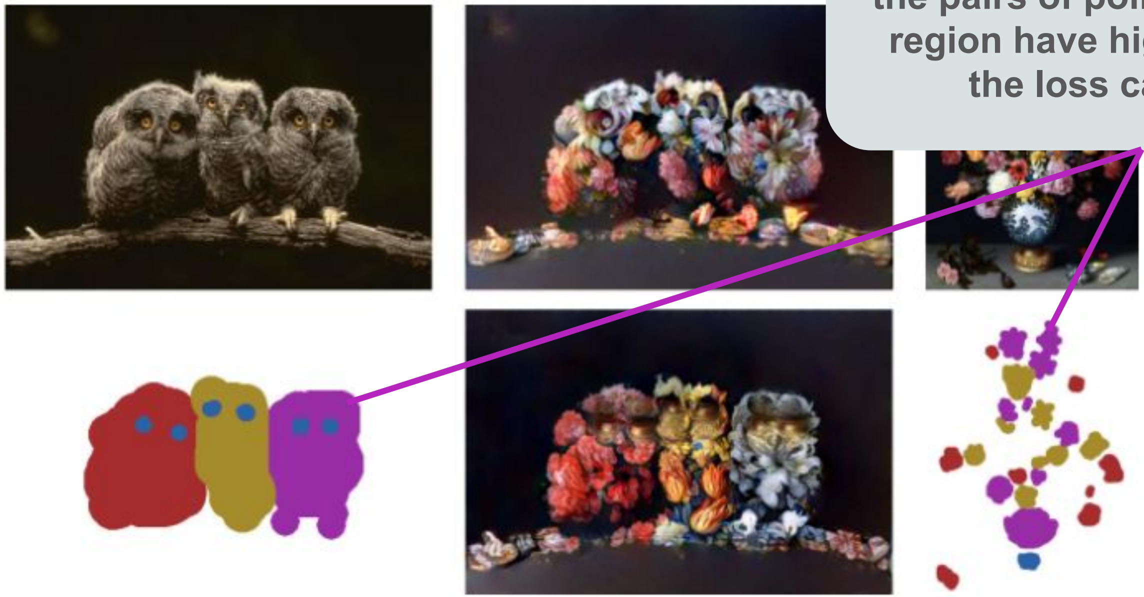
α: Content weight

The iterative process described previously is made for 4 different scales using the (upscaled) output of the previous scale as input, halving the content weight (a) for the next scale



Mask specific areas to have the same style:

Mask specific areas to have the same style:



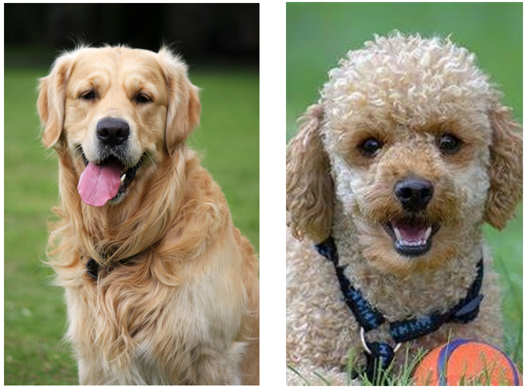
Control is enforced by making the pairs of points in the same region have higher weight in the loss calculation

EXPERIMENTS AND RELATED WORK

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Regimes :

Paired

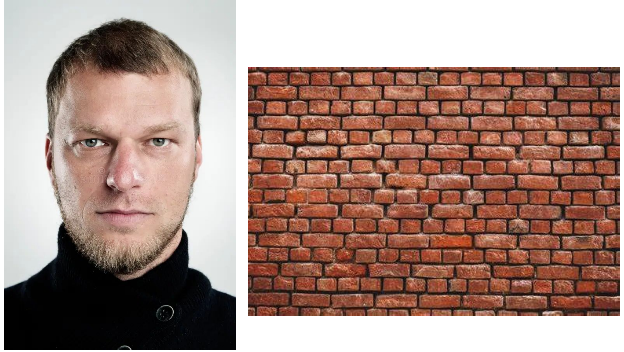


Content image = style image

Unpaired

Content image ≠ style image

Texture

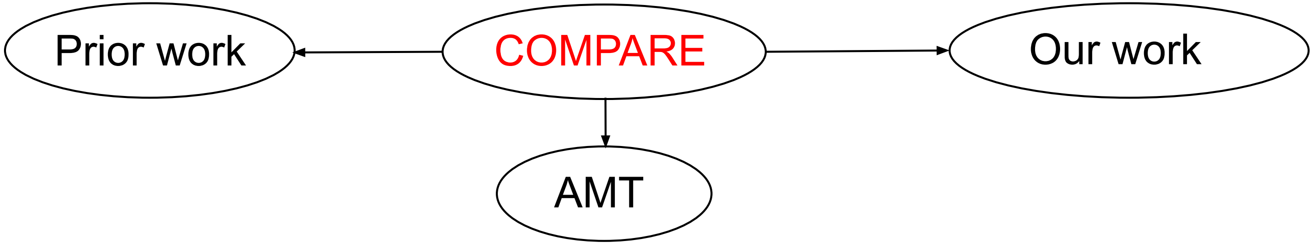


Content: Face photography
Style: Homogeneous texture

*30 style/content pairings (total of 90)

CHAPITRE 4 : EXPERIMENTS

4.1 Large-Scale human evaluation



An example of worker’s interfaces:

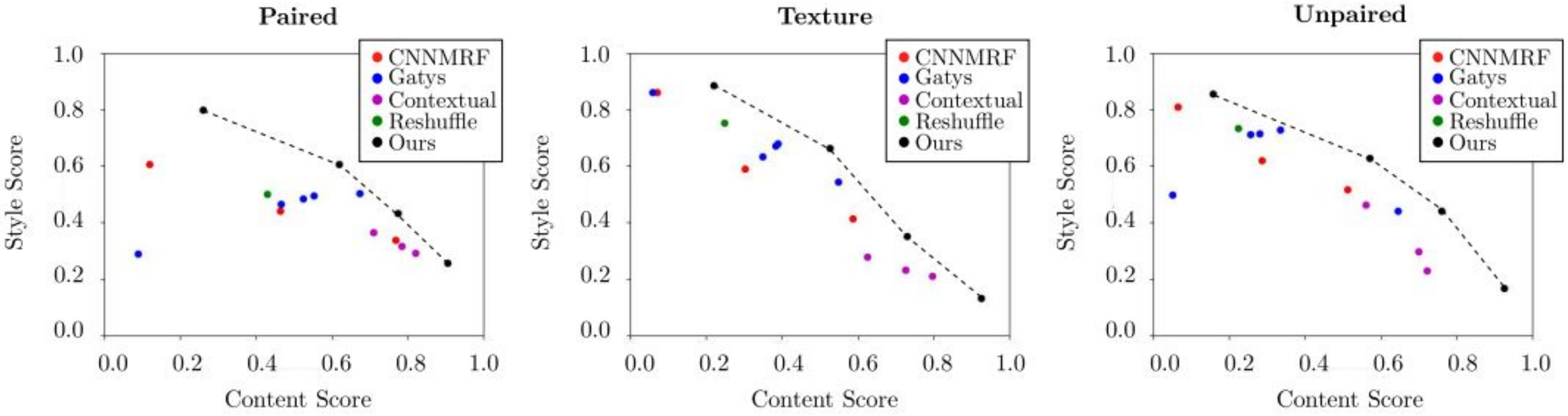
Evaluate whether Image A or Image B has more similar **style** to Image C

The style of Image C is most similar to…

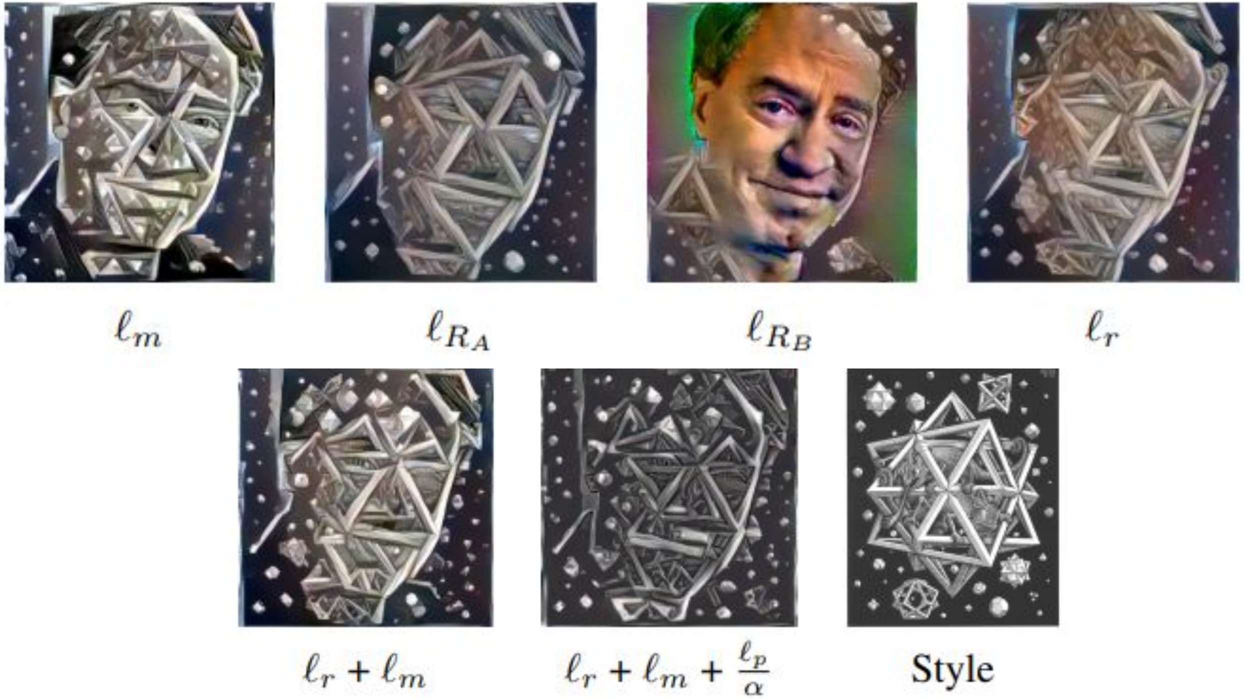
a) Image A
b) Image B
c) **Equally** A and B
d) **Neither** A or B

We test 3 sets of hyper-parameters: Content weight (high and low stylization)

The style score is defined analogously :



Effect of different terms of our style loss



4.3. Relaxed EMD Approximation Quality

$$\frac{REMD(A,B)}{EMD(A,B} \leq 1$$

4.4. Timing results

Image size	64	128	256	512	1024
Ours	20	38	60	95	154
Gatys	8	10	14	33	116
CNNMRF	3	8	27	117	X
Contextual	13	40	189	277	X
Reshuffle	-	-	-	*69**	-

CONCLUSIONS AND FUTURE WORK

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Conclusions

- Our algorithm demonstrates superior performance in style transfer compared to prior methods.
- We emphasize the importance of style-similarity losses for enhancing stylization quality.
- The simplicity and effectiveness of our earth movers distance approximation highlight its potential in style transfer applications.

Future work

- Further exploration of more accurate approximations for the earth movers distance.
- Improvement of algorithm speed through the incorporation of feed-forward style transfer methods using our proposed objective function.

QUESTIONS?

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Alex SZAPIRO

How does this method differ from Multimodal Style Transfer via Graph Cuts (Zhang et al., 2019) ?

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Pierre Monot

If the user control is also using a REMD loss, why not incorporate it in the content loss?

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Yosr DRIRA

The paper has mentioned cosine distance as a self similarity descriptor. Are there other descriptors that you could have used ?

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom