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

Texture synthesis
Using a Pseudo Optimizer (PO) and Optimal Transport (OT)

Group 4:
COSSIO Guillermo
SZAPIRO Alex
BECKER Gonzalo

Index

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

Fast Texture Synthesis via Pseudo Optimizer (PO)

Wu Shi, Yu Qiao

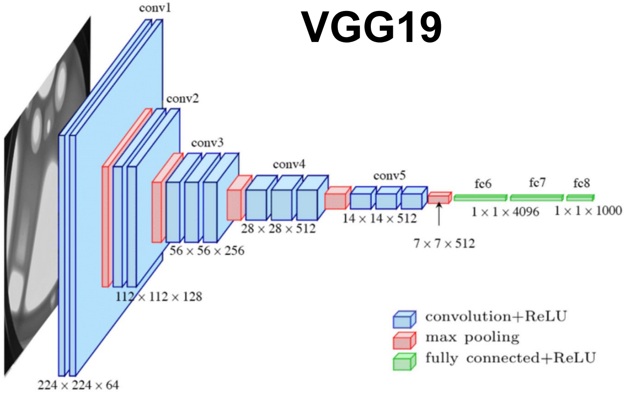
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5498-5507

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

Pseudo Optimizer (PO)

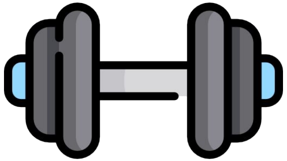
Compatibility problems

VGG19



?

Weights
(Correct format)



?

Authors

Tensorpack

NVIDIA
CUDA 9.0

GPU (8 GB)

OpenCV

Tensorflow

Pseudo Optimizer (PO)

Running the algorithm

Original code used:

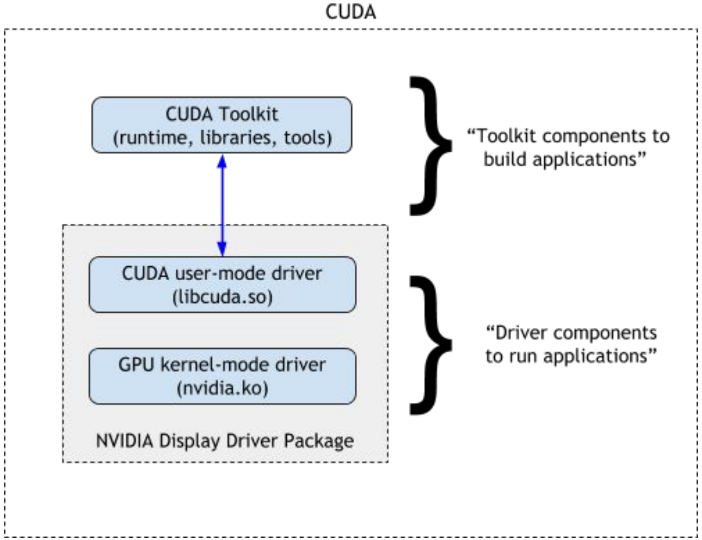
- Python 3.6
- Tensorflow 1.12
- Tensorpack 0.9.8
- Opencv 3.4.2.17

Environment:

- conda create -n syntex python=3.6 cudatoolkit=9.0
- conda install tensorflow-gpu=1.12
- pip install tensorpack==0.9.8
- pip install opencv-python==3.4.2.17

Main problems with the repo:

- VGG pretrained file was missing!
- ReadMe's were not clear at all



2. Results

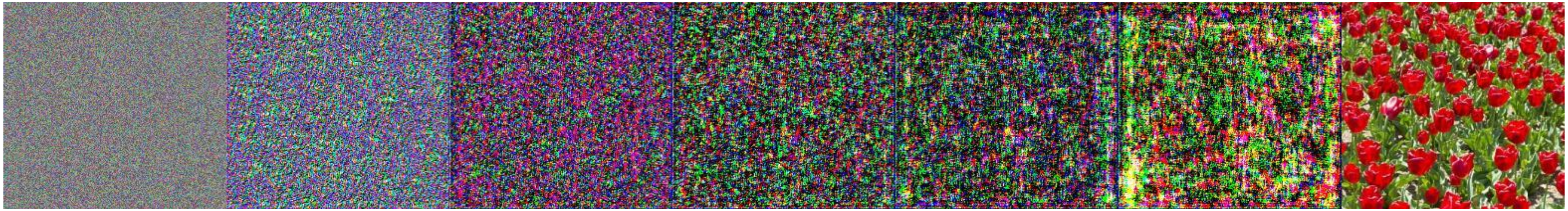
IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Results

Effect of training length on results

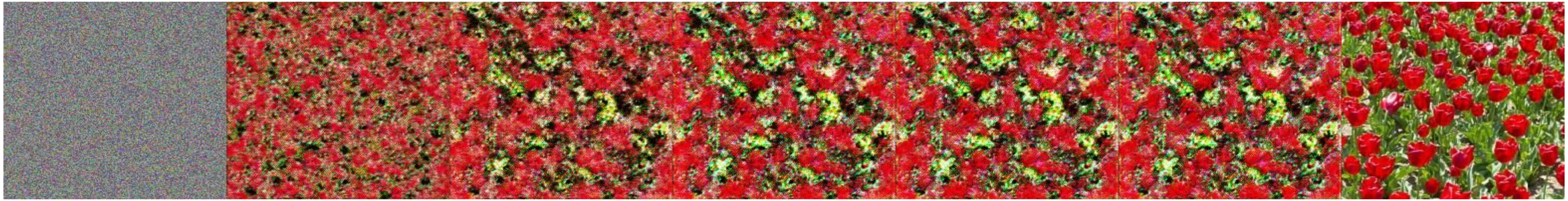
Train size = 38 (flower images) | Epochs = 10 | Steps = 100

input $x^{[0]}$ | $x^{[1]}$ | $x^{[2]}$ | $x^{[3]}$ | $x^{[4]}$ | $x^{[5]}$ | Target

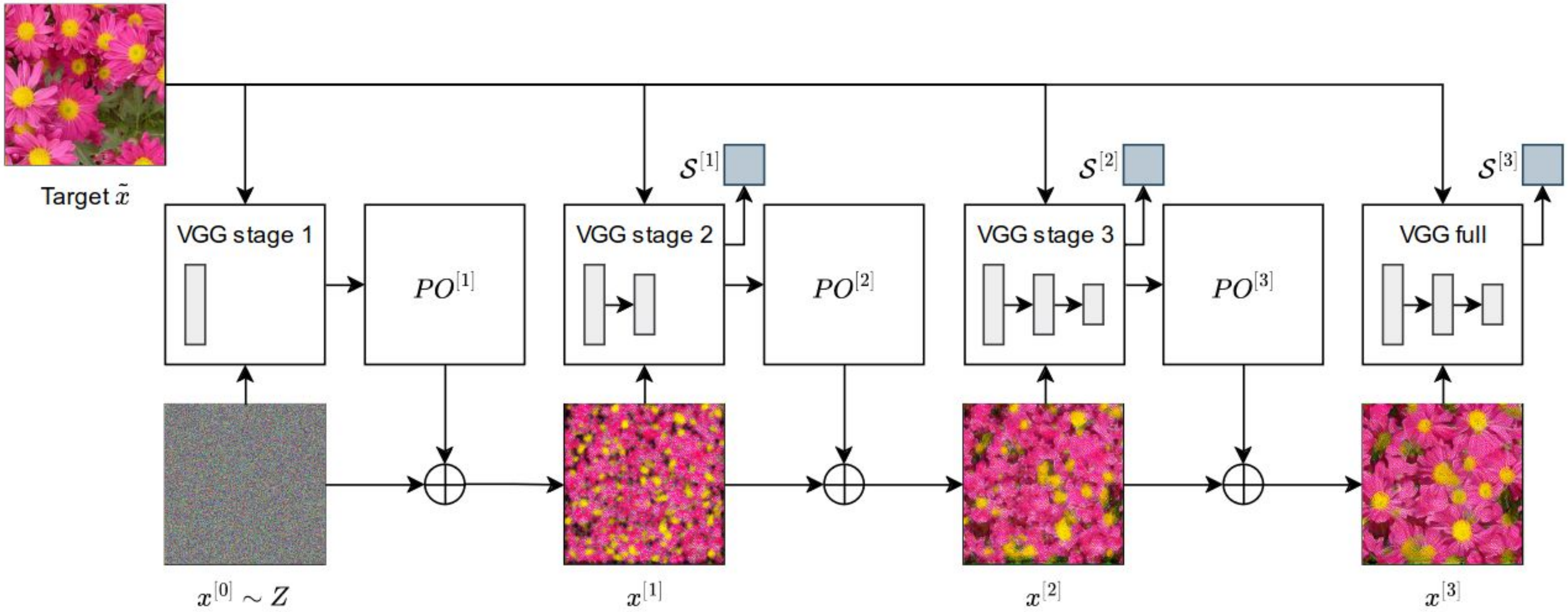


Train size = 38 (flower images) | Epochs = 100 | Steps = 100

input $x^{[0]}$ | $x^{[1]}$ | $x^{[2]}$ | $x^{[3]}$ | $x^{[4]}$ | $x^{[5]}$ | Target

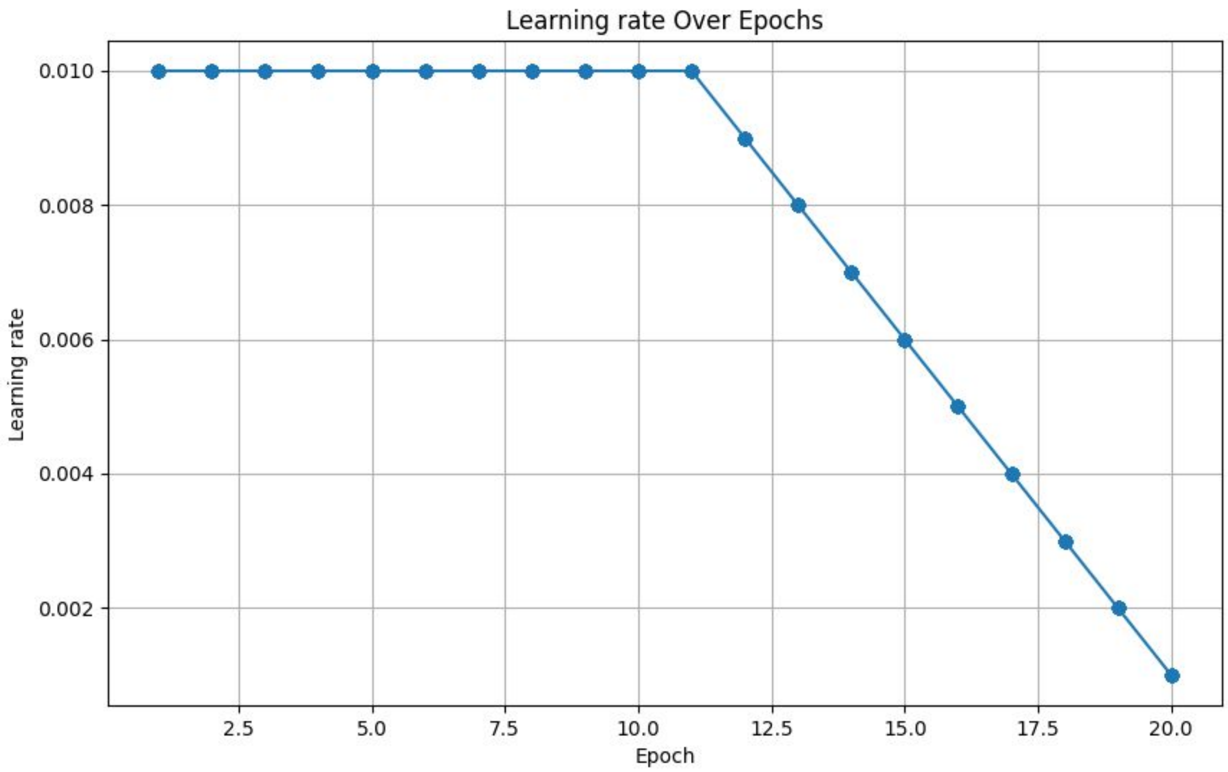


Progressive Pseudo Optimizer Architecture



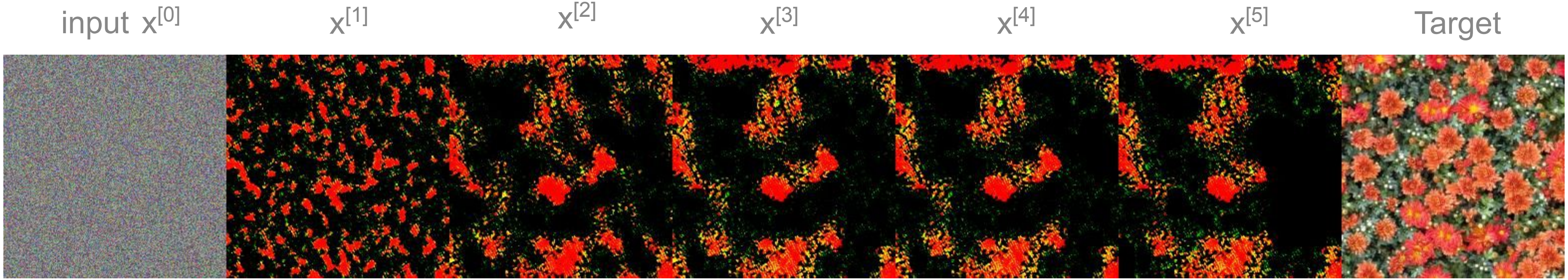
Learning Rate

20 epochs, 100 steps per epoch, initial LR = 0.01

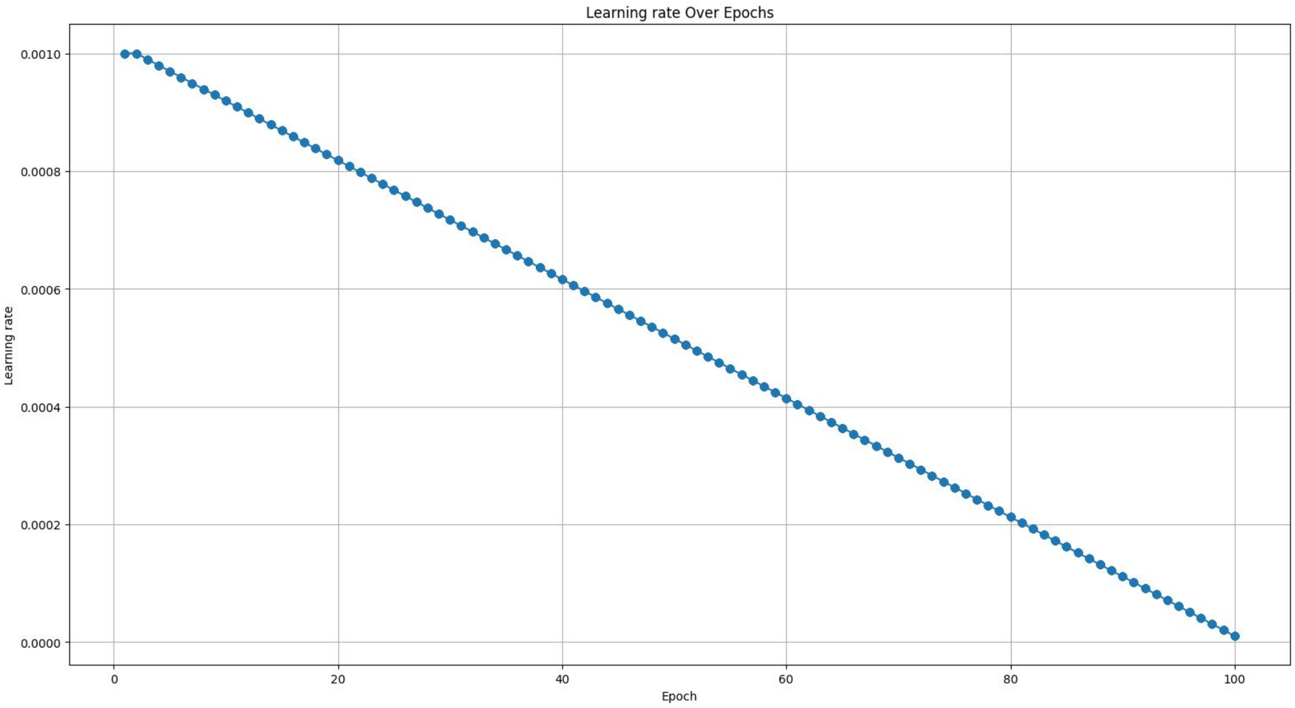


Results

100 epochs, 300 steps per epoch, initial LR = 1E-3

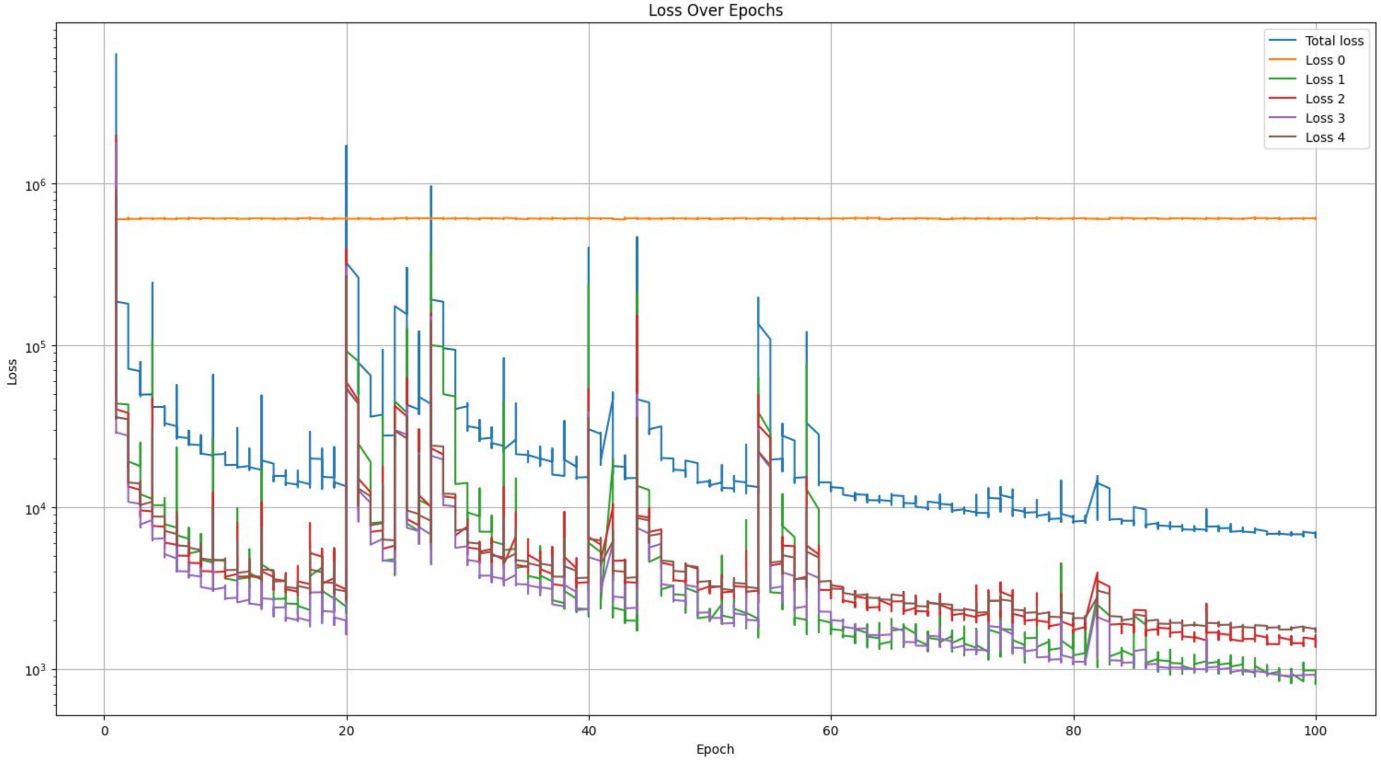


Learning Rate



Loss evolution

As an average of all layers



Results

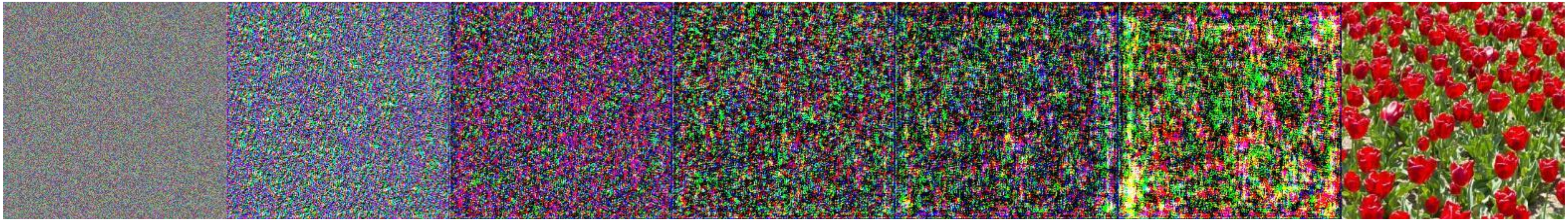
100 epochs, 300 steps per epoch, initial LR = 1E-3

Improved Progressive Pseudo Optimizer

Results

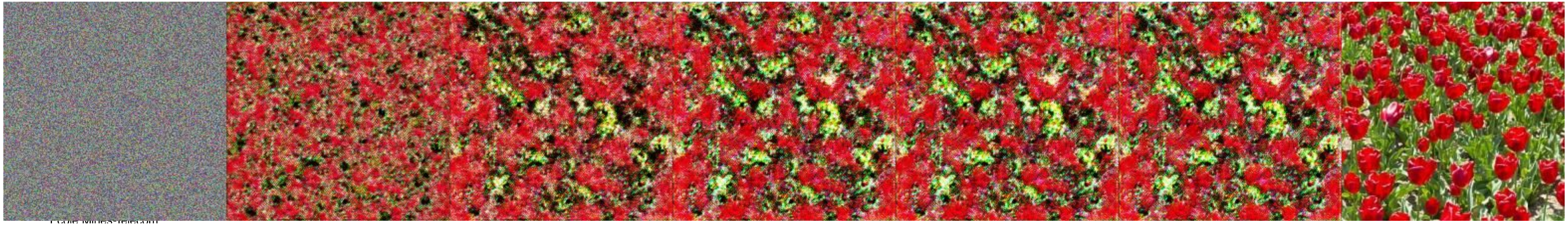
<u>Before</u>: Train size = 38 (flower images) | Epochs = 10 | Steps = 100

Input $x^{[0]}$ | $x^{[1]}$ | $x^{[2]}$ | $x^{[3]}$ | $x^{[4]}$ | $x^{[5]}$ | Target



<u>After</u>: Train size = 38 (flower images) | Epochs = 100 | Steps = 500

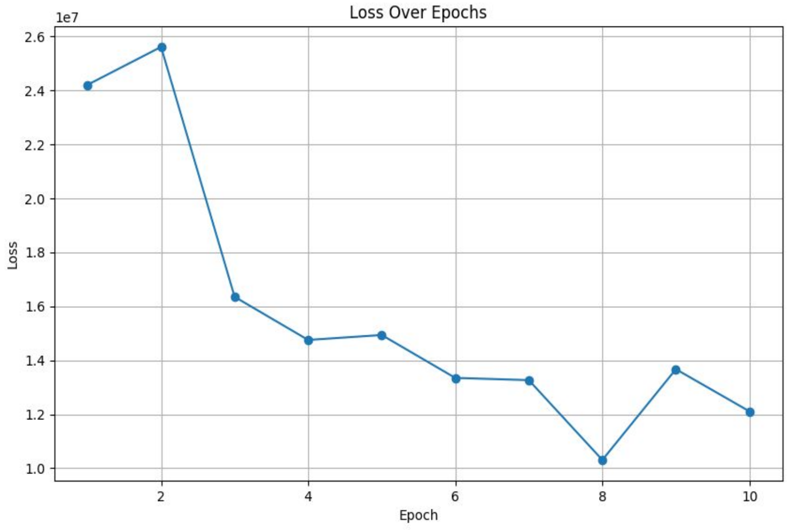
Input $x^{[0]}$ | $x^{[1]}$ | $x^{[2]}$ | $x^{[3]}$ | $x^{[4]}$ | $x^{[5]}$ | Target



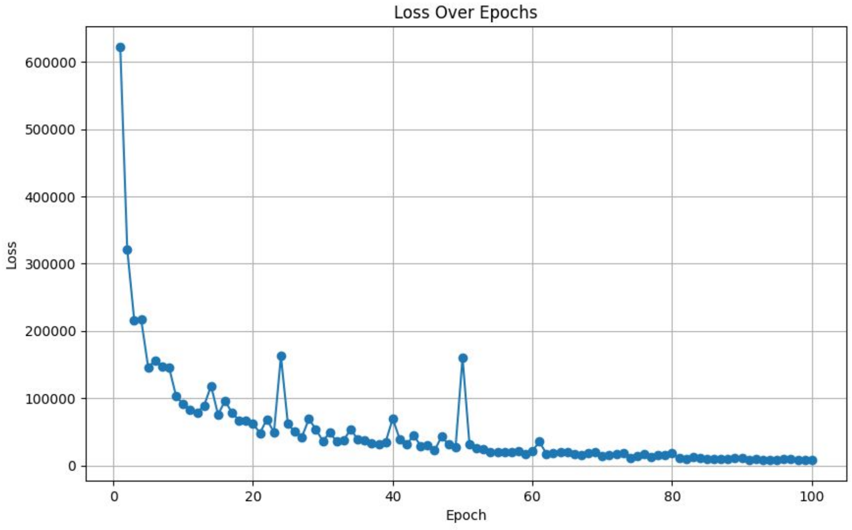
Improved Progressive Pseudo Optimizer

Loss

<u>Before</u>



<u>After</u>

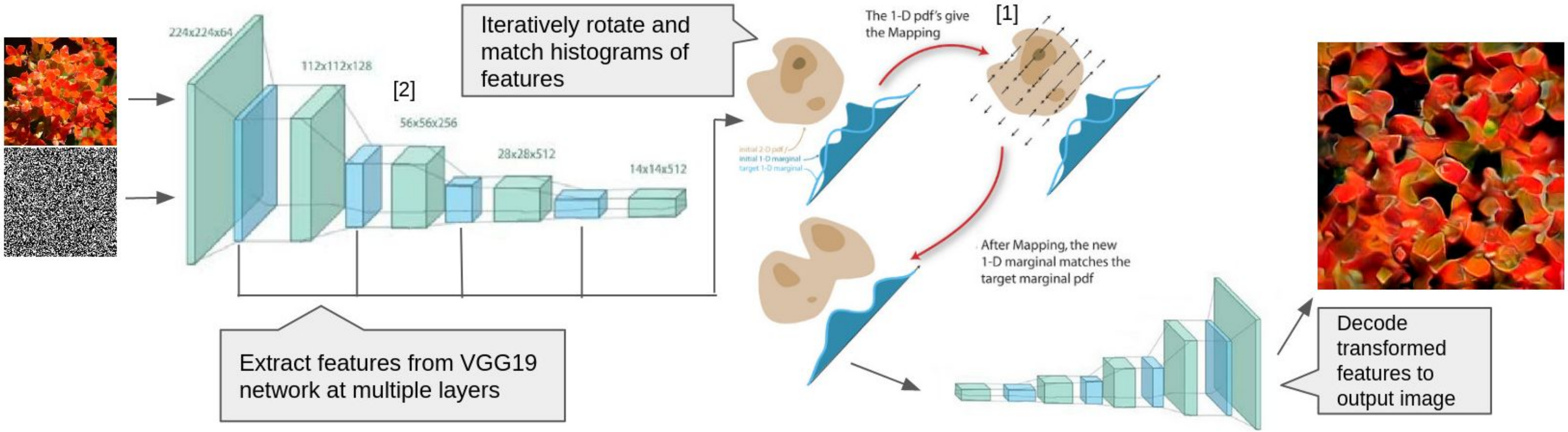


Optimal Transport

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

Optimal Transport

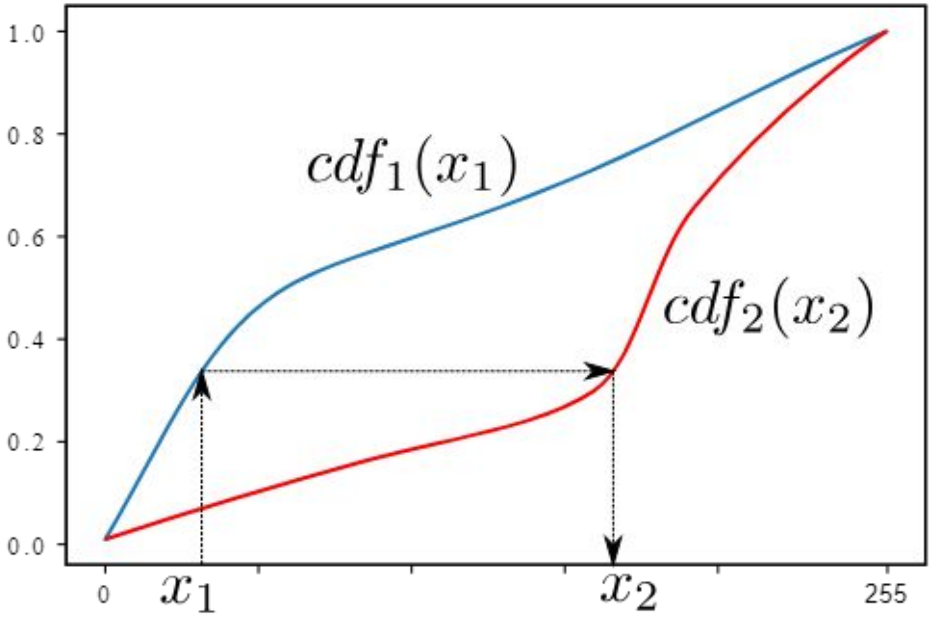
No-training method for texture synthesis



Eric Risser, *"Optimal Textures: Fast and Robust Texture Synthesis and Style Transfer through Optimal Transport"*, arXiv:2010.14702 [cs.GR]

Optimal Transport

Histogram Matching



Accelerate this process by applying

- **PCA,**
- **Cholesky decomposition,**
- **or Symmetric eigenvalues**

on the covariance matrix of the histogram.

Optimal Transport

Results

Symmetric strategy	PCA strategy	Cholesky Decomposition strategy	Cumulative Distribution Function strategy
			
Time: 38.22 s	41.22 s	23.50 s	262.79 s

Not adapted to real-time applications!

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

Optimal Transport
Conclusions

- Much like the original method (Gatys et al.) this algorithm doesn’t require training, but is far too slow for real-time applications.
- Different statistics about the internal features of VGG19 also provide good results!

3. Conclusions

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

Conclusions

Results of ProPO and OT

*Image taken from the paper, not our implementation

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

Conclusions
Challenging images

The images that were challenging for the ProPO (geometrical regularities) did not get better results with OT.

ProPO*

Target

OT

*Image taken from the paper, not our implementation

Thanks for your attention!
Any questions?

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

1. Old

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

Texture synthesis

What is the problem targeted in this paper?

Address the challenge of accelerating **texture synthesis** while maintaining high quality and diversity in the generated textures.

What is texture synthesis?

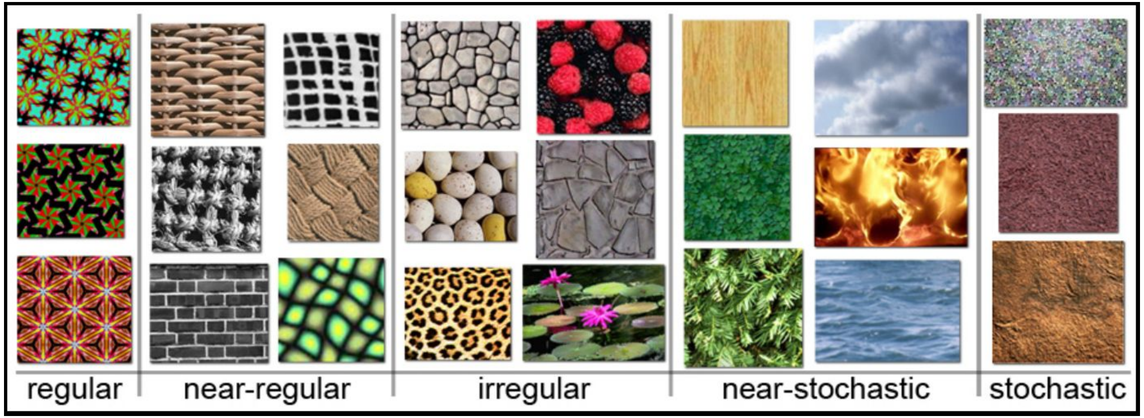
Process of generating new *textures* given an original one.

↓ i.e.

Creating images/patterns

↓

Similar visual properties, statistical characteristics, and structural features as the original texture.

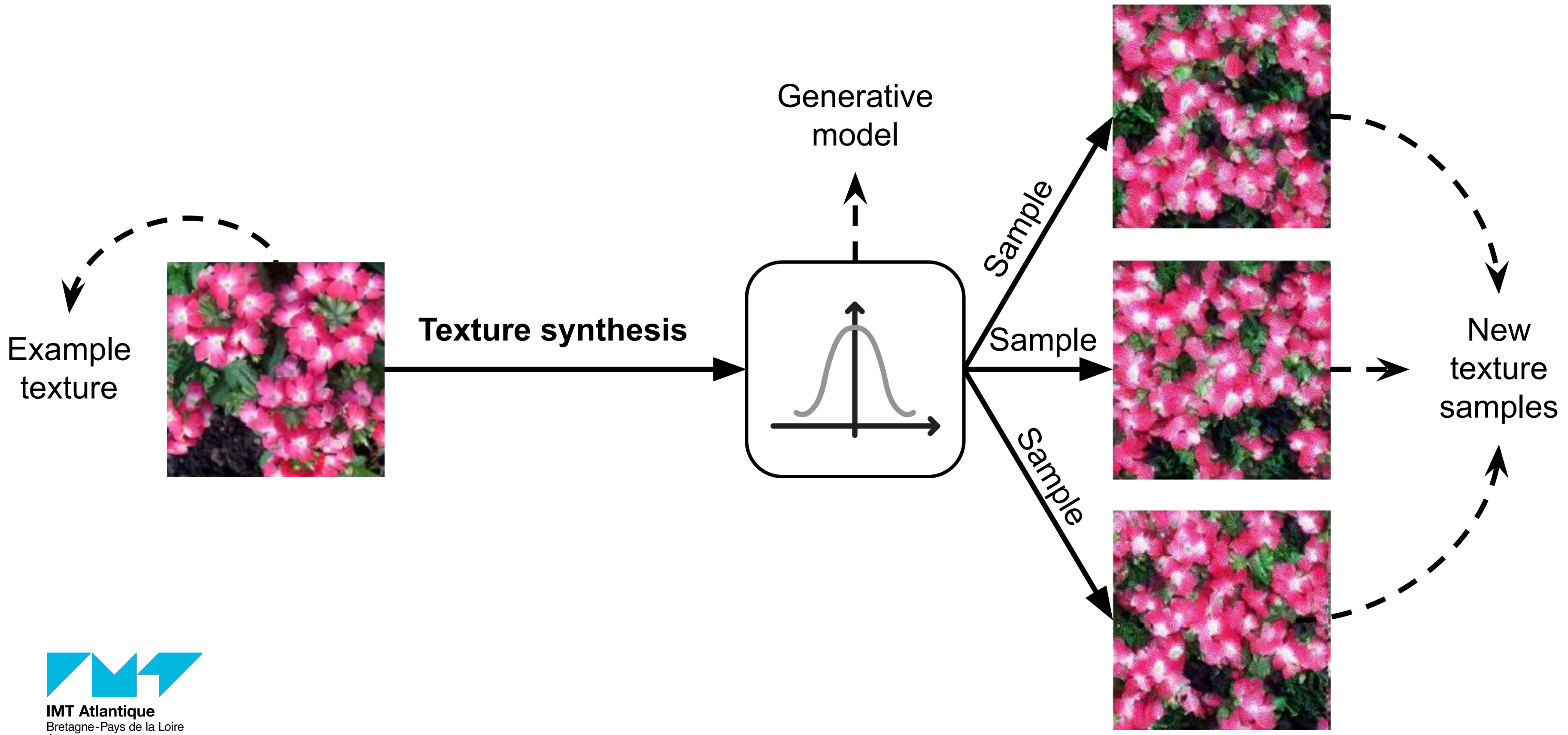


"Near-regular Texture Analysis and Manipulation." Yanxi Liu, Wen-Chieh Lin, and James Hays. SIGGRAPH 2004

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

Texture synthesis

What is the problem targeted in this paper?



Minimization problem

Defining the objective function

Part of fig. 6 of Shi & Qiao (2020), showing examples of texture synthesis using the algorithm in Gatys et al. (2015)

From an image $\tilde{x}$ we want to get a sample $\hat{x}$ from the set of images that share structural similarities with $\tilde{x}$.

How do we define an appropriate objective function?

Minimization problem

Defining the objective function

The objective function

Function that quantifies the difference between:

$$\arg\min_{x \in \mathcal{X}} \mathcal{L}_{tex}(x, \tilde{x}; [L]) = \sum_{l=1}^{L} \left\| G^{(l)}(x) - G^{(l)}(\tilde{x}) \right\|_2^2,$$

- the synthesized texture
- the target texture

The goal is to minimize this function so to generate a synthesized texture that closely matches the properties of the target texture

Texture Loss

3.1. Texture Synthesis using CNN

The authors of [5] reduce texture synthesis to the problem of sampling from the set of images that match the spatial summary statistics of the example texture ima
They use the VGG19 network [18], a convolutional n
ral network trained on object classification, to extrac
set of powerful descriptive feature activations: $\{F^{(l)}$
$\mathbb{R}^{N^{(l)} \times C^{(l)}}\}_{l=1}^{L}$, where (l) is the index of layer, $N^{(l)}$ is
spatial dimension and $C^{(l)}$ is the number of channel
layer (l). The summary statistics are defined by the c
relations, *i.e.* the Gram matrix[1] $G^{(l)} \in \mathbb{R}^{C^{(l)} \times C^{(l)}}$, betw
the responses of different features:

$$G_{ij}^{(l)} = \frac{1}{N^{(l)}} \sum_{k \in [N^{(l)}]} F_{k,i}^{(l)} F_{k,j}^{(l)}, \quad i, j \in [C^{(l)}]. \qquad (1)$$

To synthesize a new texture given the example image $\tilde{x}$, they pose the generation process as an optimization problem by solving

$$\arg\min_{x \in \mathcal{X}} \mathcal{L}_{tex}(x, \tilde{x}; [L]) = \sum_{l=1}^{L} \left\| G^{(l)}(x) - G^{(l)}(\tilde{x}) \right\|_2^2, \qquad (2)$$

where $\mathcal{X}$ is the image space and $[L]$ denotes the set of layers involved in the calculation. The objective function $\mathcal{L}_{tex}$ is usually named as texture loss in the related works. In

2. Key concepts

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

Key Concepts
Texture synthesis

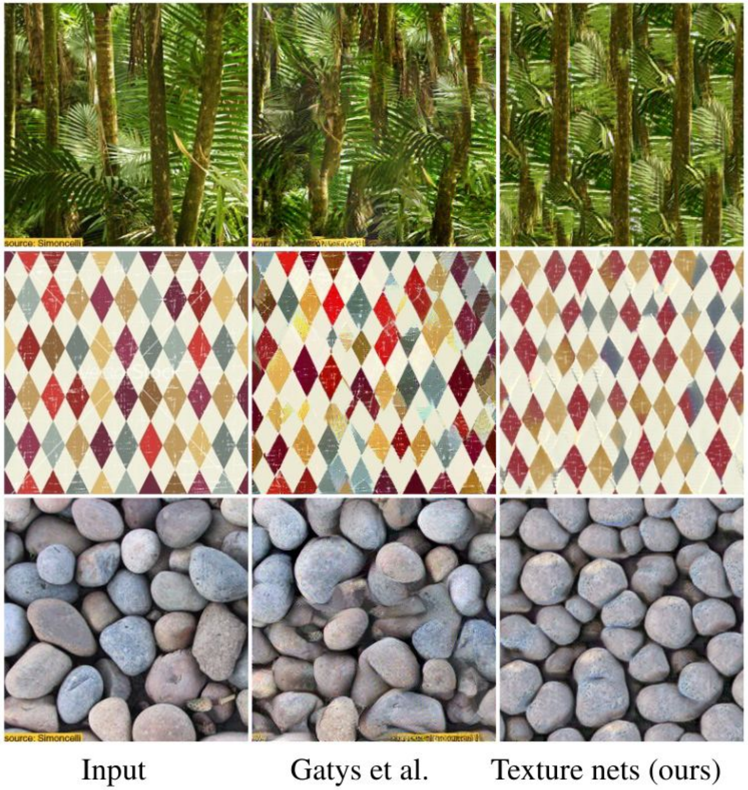
Part of fig. 6 of Shi & Qiao (2020), showing examples of texture synthesis using the algorithm in Gatys et al. (2015)

From an image $\tilde{x}$ we want to get a sample x from the set of images that share structural similarities with $\tilde{x}$.

The question arises: **how do we define an appropriate cost function?**

Metrics have to consider entire structures rather than individual pixel values.

Key Concepts
Texture synthesis



Part of fig. 1 of Ulyanov et al. (2016). Note how pictures with similar textures can be very different pixel-wise.

Key Concepts
Features

We can use another neural network to get appropriate features!

A network trained for classification generates internal features that describe the structure of the input image.

Comparing these features allows us to get an adequate loss function!

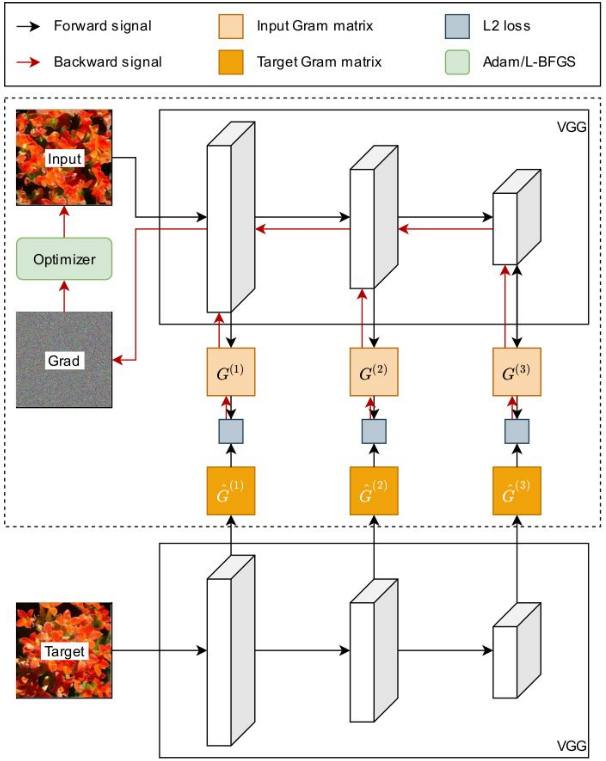
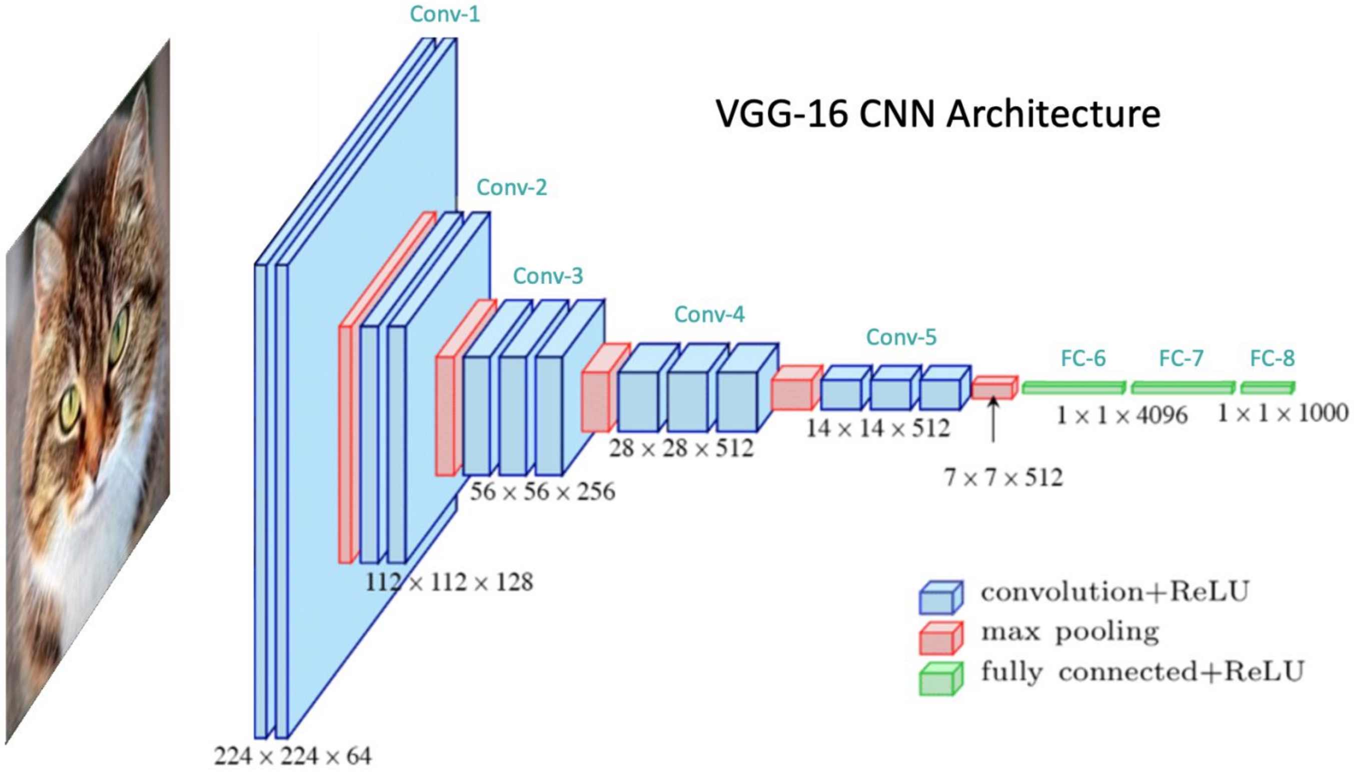


Fig. 2 of Shi & Qiao (2020), showing how features are extracted from the internal activations of a CNN (in this case VGG19).

Key Concepts
Features



Key Concepts
Gram matrix

What exactly are the features?

“**Summary statistics** are used to summarize a set of observations, in order to communicate the largest amount of information as simply as possible.”

Correlation between the pixels values is important.

How to represent the features conveniently? **Gram matrices.**

$$G_{ij}^{(l)} = \frac{1}{N^{(l)}} \sum_{k \in [N^{(l)}]} F_{k,i}^{(l)} F_{k,j}^{(l)}, \quad i, j \in [C^{(l)}]$$

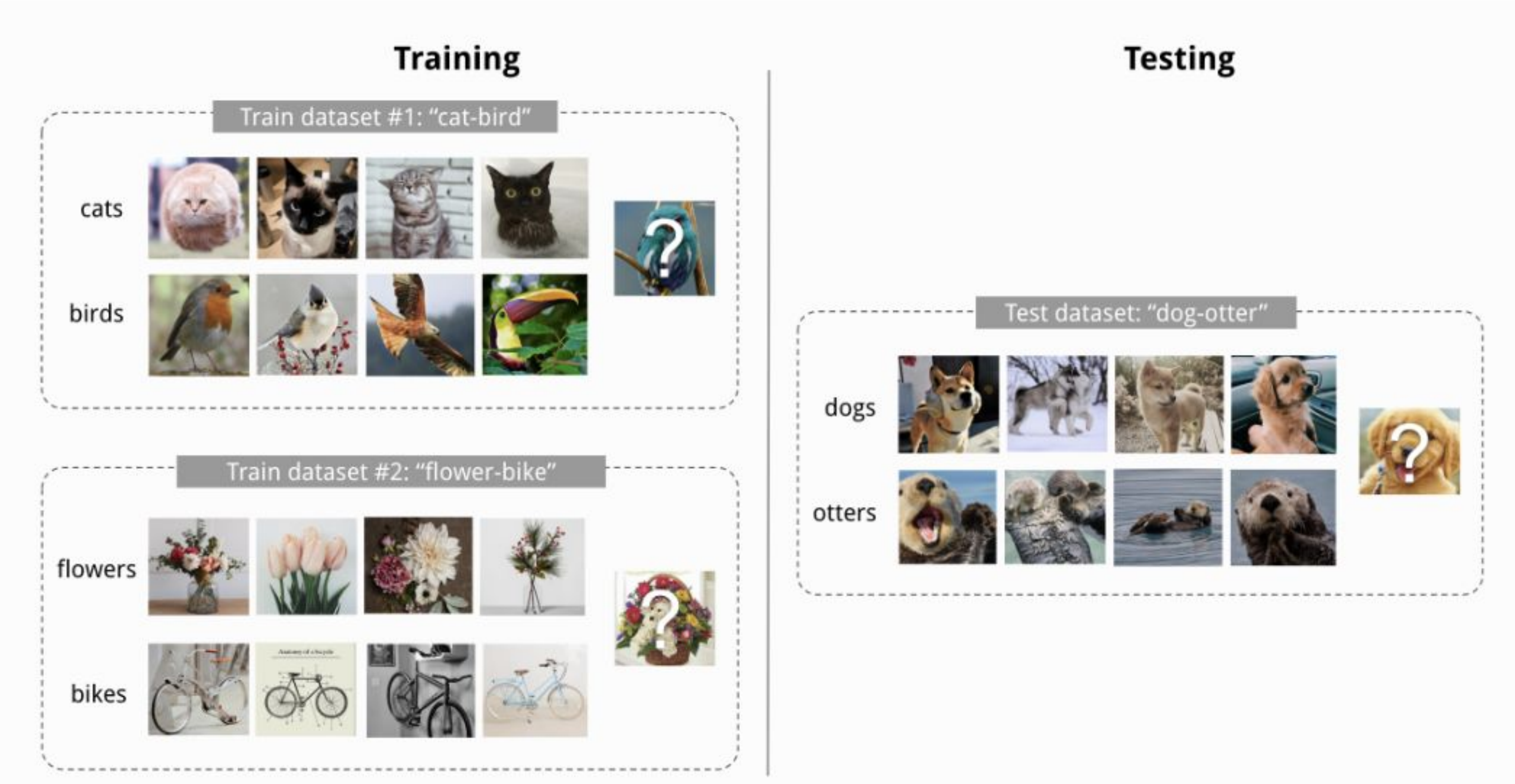
Key Concepts

Meta learning: “Learning to learn”

Use data from other previously seen tasks to **learn how to learn.**
Then learn the new tasks more efficiently with small amount of data.

Instead of learning from a distribution of images, we learn from a distribution of tasks.

Key Concepts
Meta learning: example



Meta-learning example: few shot learning.
Credit: https://lilianweng.github.io/posts/2018-11-30-meta-learning/

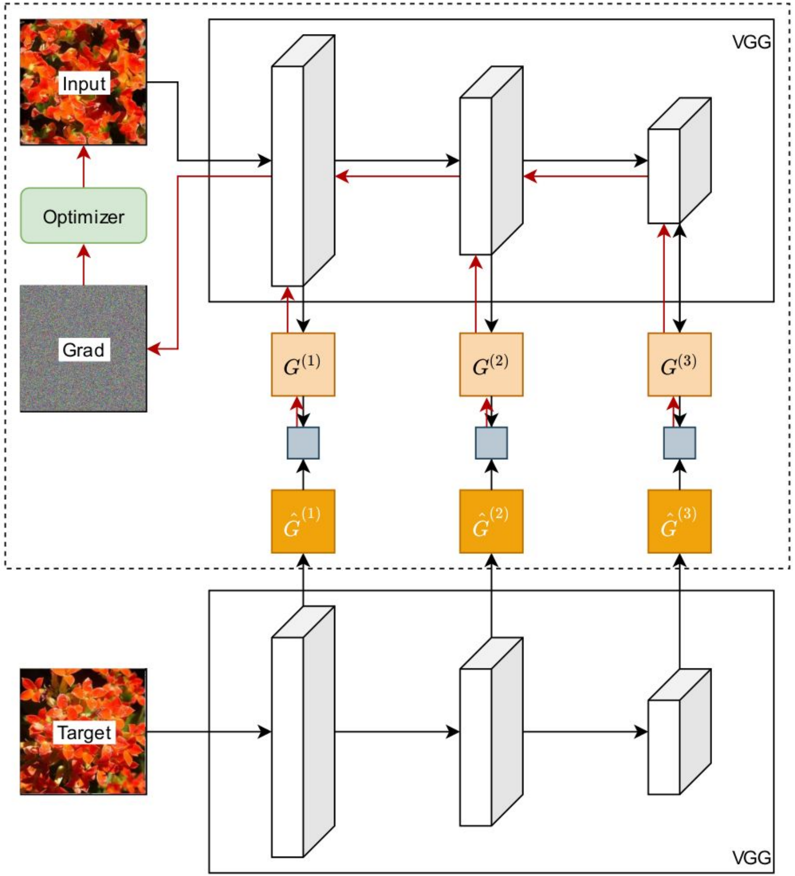
3. Method

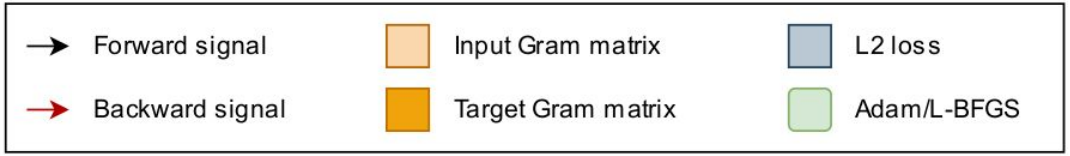
IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Original method

The method developed by Gatys et al. in 2015 generates textures in an iterative manner

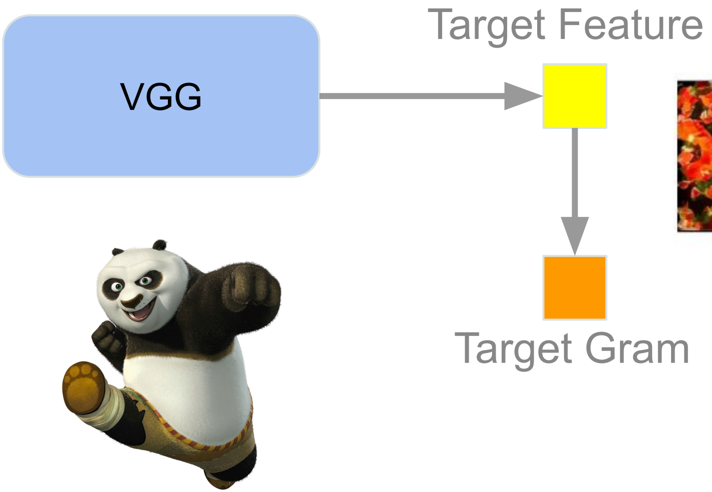
- Good quality
- Scalable
- Very slow to compute

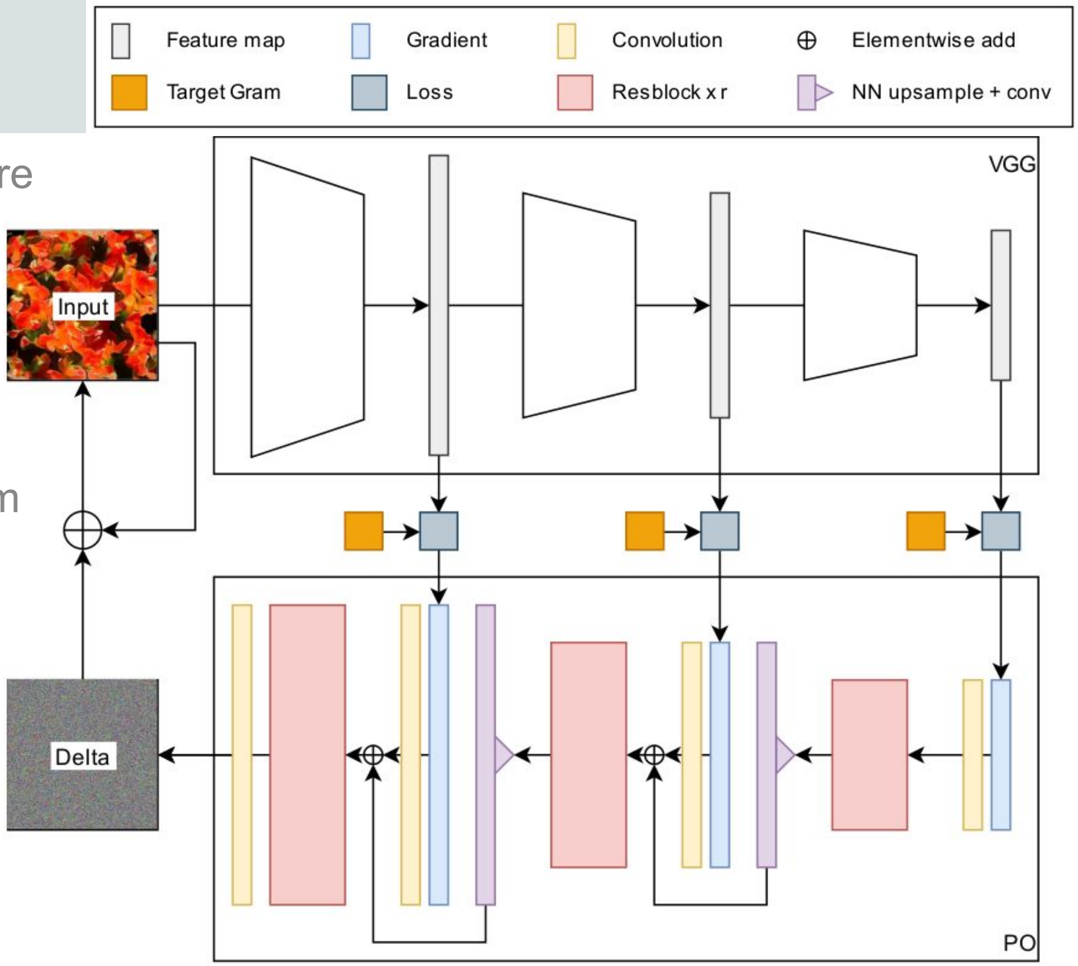




Pseudo Optimizer (PO)

Simulate the optimization process



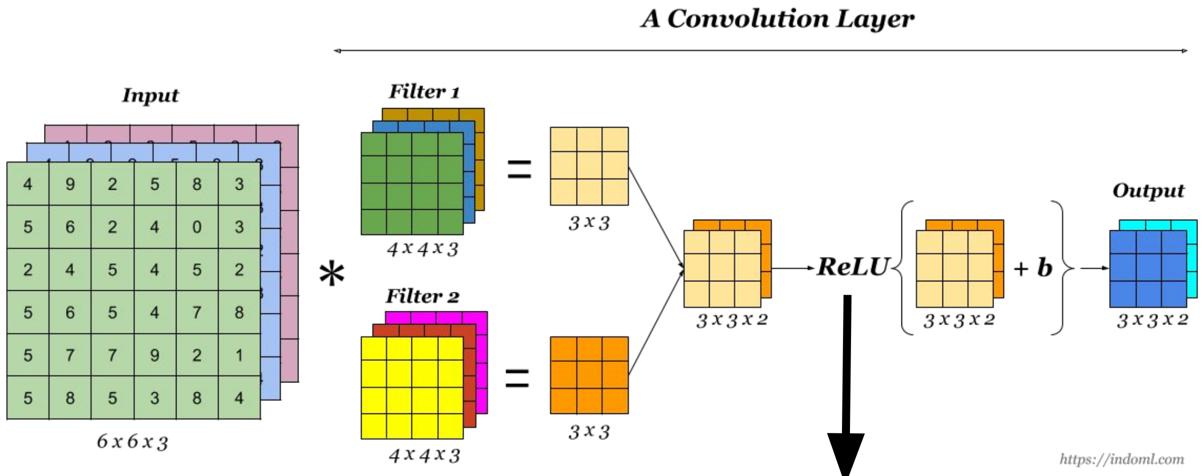


$$\Delta x = PO\left(\left\{\frac{\partial \mathcal{L}_{tex}(x, \tilde{x}; [L])}{\partial F^{(l)}(x)}\right\}_{l=1}^{L}\right)$$

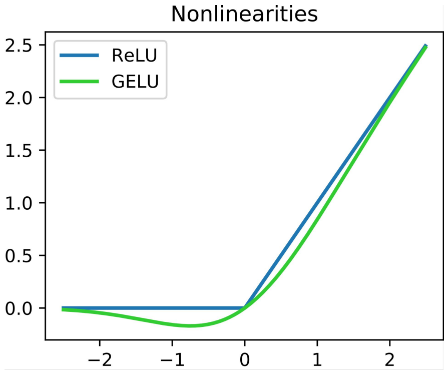
IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Pseudo Optimizer (PO)

Convolutional layer



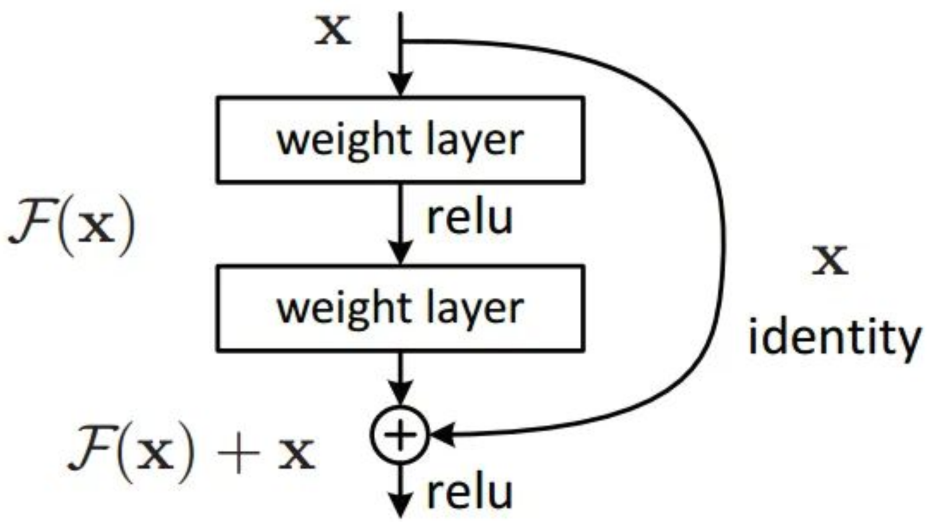
Activation Function



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

Pseudo Optimizer (PO)

Residual block



Resblocks are used in neural networks to facilitate the training of deeper architectures by alleviating the vanishing gradient problem and enabling easier information flow through skip connections.

Comparison
Between both methods

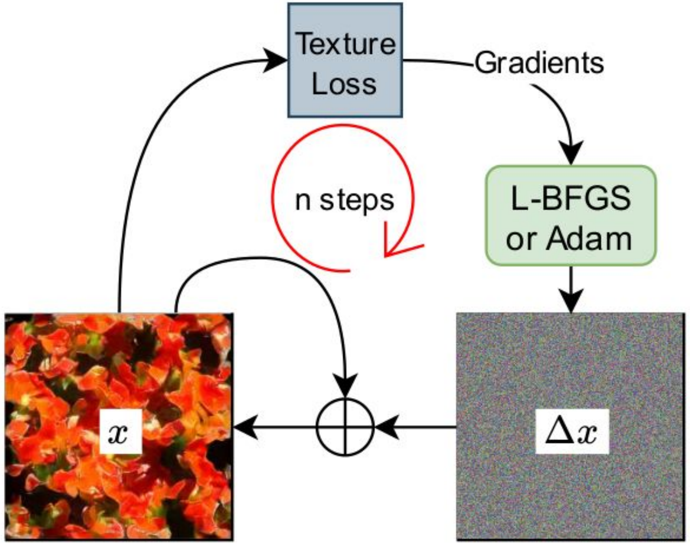


Diagram of the algorithm by Gatys et al. (2015)

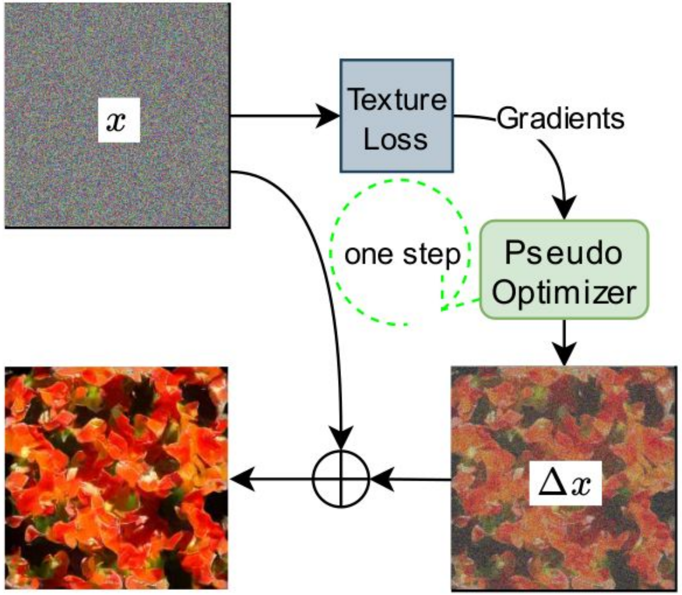


Diagram of the proposed algorithm by Shi and Qiao (2020)

Pseudo Optimizer (PO)

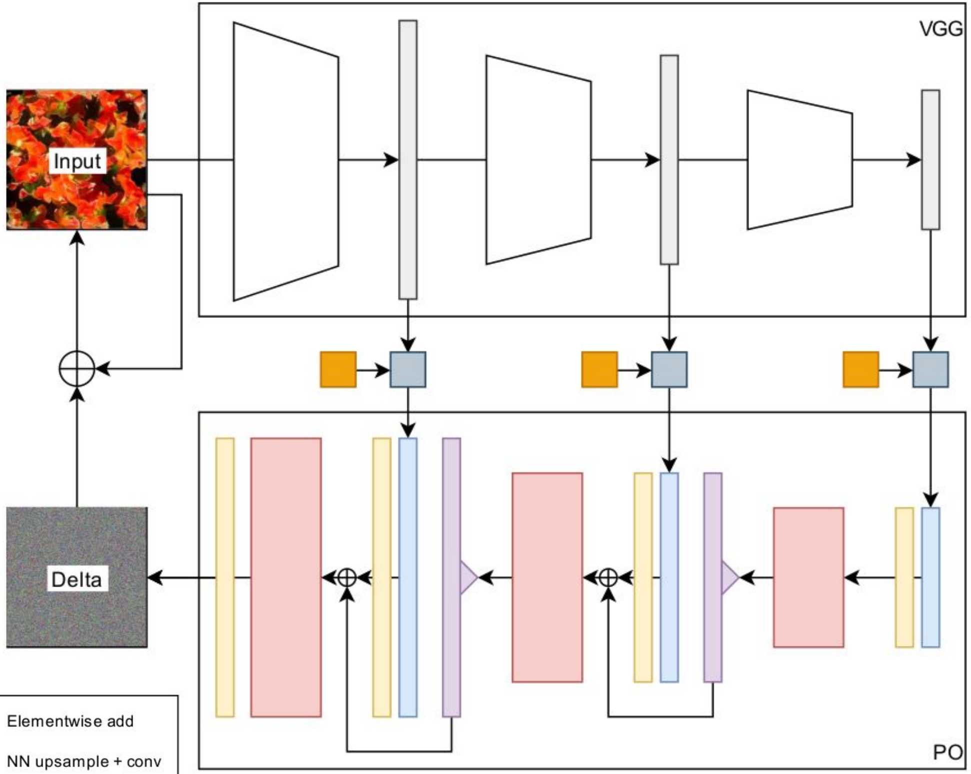
Simulate the optimization process for real-time applications

Simulate the optimization process

Only 1 target image is used for training

Not adaptive!

$$\Delta x = PO\left(\left\{\frac{\partial \mathcal{L}_{tex}(x,\tilde{x};[L])}{\partial F^{(l)}(x)}\right\}_{l=1}^{L}\right)$$



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

What loss function should be used for training?

Choose a new loss function to allow for training of the pseudo optimizer

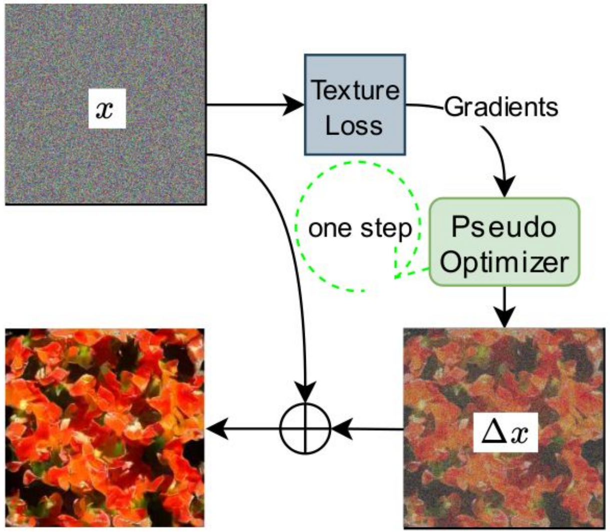
$$\mathcal{L}_{tex}(x, \tilde{x}; [L]) = \sum_{l=1}^{L} \left\| G^{(l)}(x) - G^{(l)}(\tilde{x}) \right\|_2^2$$

Cannot be computed for all ***x***

Define a new loss

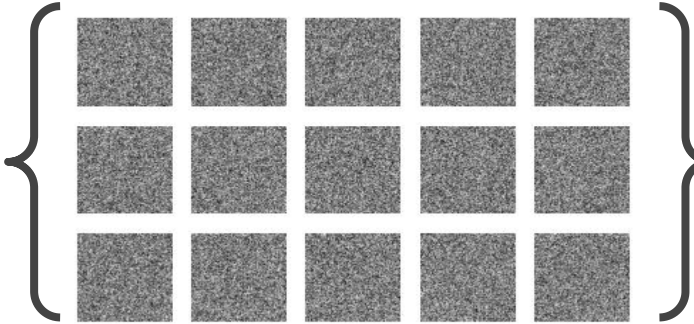
$$\mathcal{S}^{[L]}(\tilde{x}) \triangleq \mathop{\mathbb{E}}_{x \sim Z} \mathcal{L}_{tex}(x + \boxed{\Delta x}, \tilde{x}; [L])$$

$$\boxed{\Delta x = PO\left(\left\{ \frac{\partial \mathcal{L}_{tex}(x, \tilde{x}; [L])}{\partial F^{(l)}(x)} \right\}_{l=1}^{L} \right)}$$

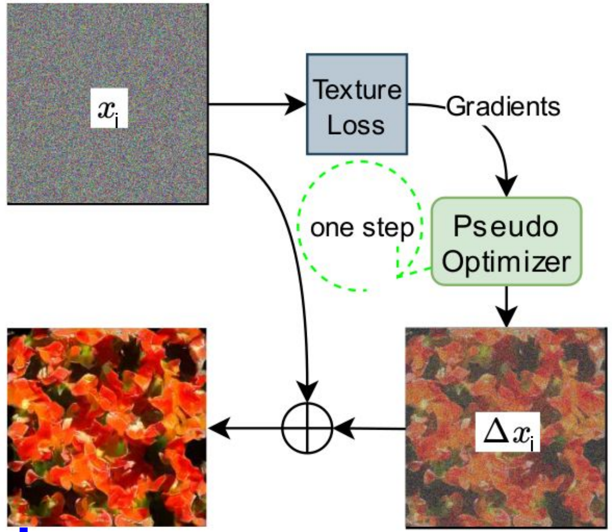


What loss function should be used for training?

Choose a new loss function to allow for training of the pseudo optimizer



Choose 1 noise input each time



$$\mathcal{S}^{[L]}(\tilde{x}) \triangleq \mathop{\mathbb{E}}_{x\sim Z} \mathcal{L}_{tex}(x+\Delta x, \tilde{x}; [L])$$

So the expectation is no more that an average of the loss for many white noise input image "seeds"

Evaluate loss for each output

$$\left\{ \mathcal{L}_{tex}(x_i + \Delta x_i, \tilde{x}; [L]) \right\}$$

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

Adaptive Pseudo Optimizer (AdaPO)

How can we make the previous algorithm adaptive?

Use a sequence of target images instead of only 1

$$\{\tilde{x}_i\}_{i=1}^{n}$$

$$\mathcal{S}_{ada}^{[L]}(\{\tilde{x}_i\}_{i=1}^{n}) \triangleq \frac{1}{n}\sum_{i=1}^{n}\boxed{\mathcal{S}^{[L]}}(\tilde{x}_i)$$

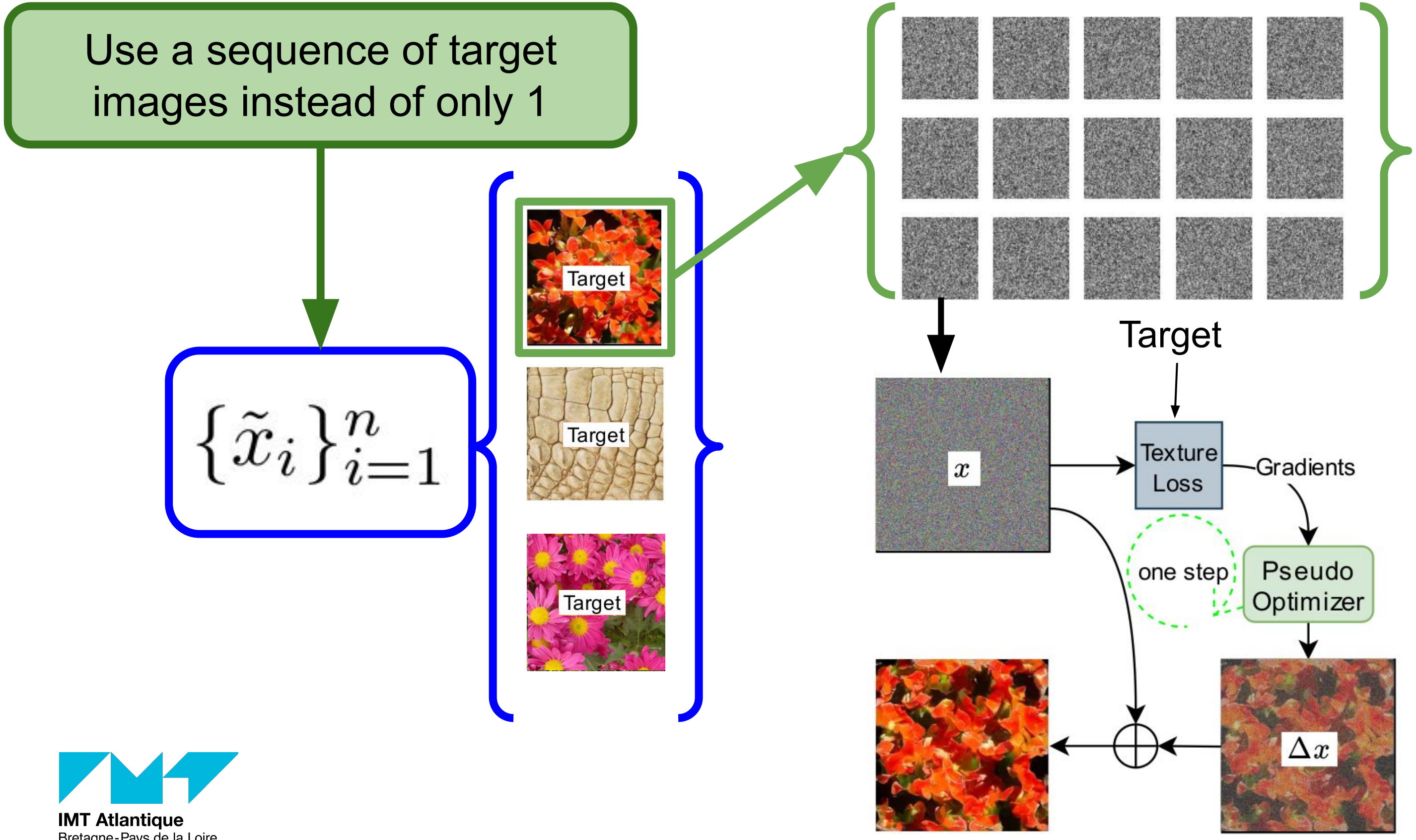
$$\mathcal{S}^{[L]}(\tilde{x}) \triangleq \mathop{\mathbb{E}}_{x\sim Z}\mathcal{L}_{tex}(x+\boxed{\Delta x}, \tilde{x}; [L])$$

$$\Delta x = PO\left(\left\{\frac{\partial \mathcal{L}_{tex}(x, \tilde{x}; [L])}{\partial F^{(l)}(x)}\right\}_{l=1}^{L}\right)$$

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

Adaptive Pseudo Optimizer (AdaPO)

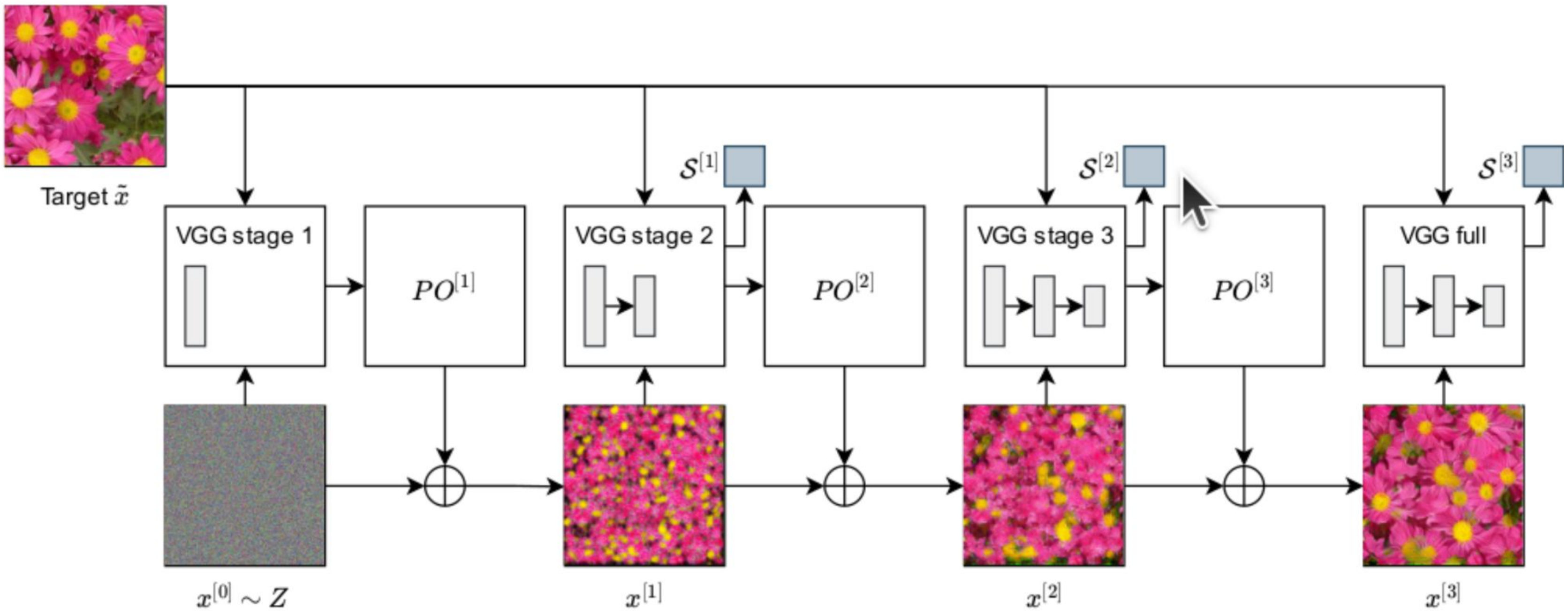
How can we make the previous algorithm adaptive?



Train the model for each target by using multiple noise input images

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

Progressive Pseudo Optimizer



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

4. Experiments & Results

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

Methods comparison

Quality & Diversity

Quality = Texture Loss
(Multi-Layer Gram Matrix)

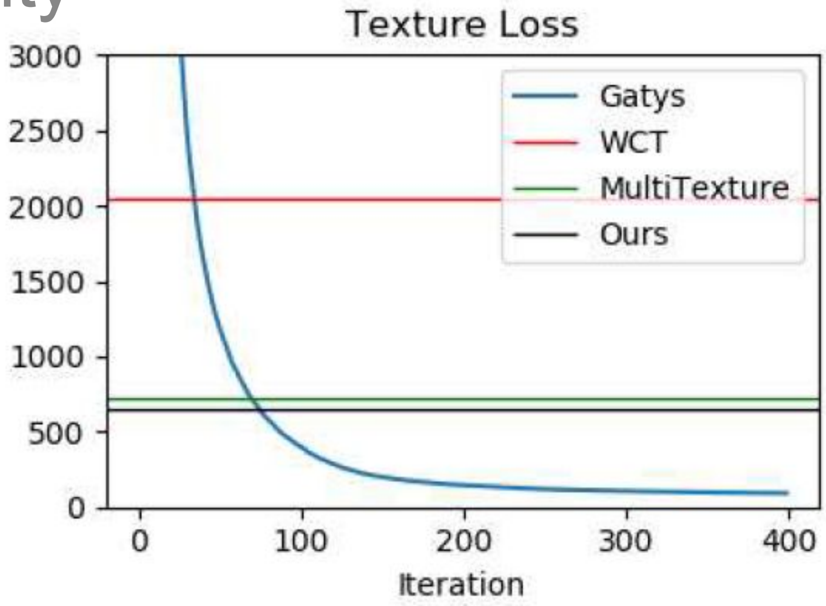
$$\mathcal{L}_{tex}(x, \tilde{x}; [L]) = \sum_{l=1}^{L} \left\| G^{(l)}(x) - G^{(l)}(\tilde{x}) \right\|_2^2$$

Diversity = Diversity Loss
(Single-Layer Feature Map)

$$\mathcal{L}_{div}(\{x_i\}_{i=1}^{b}) = \mathop{\mathbb{E}}_{i \neq j} \left\| F^{(div)}(x_i) - F^{(div)}(x_j) \right\|_1$$

Floral (47) — Preliminary experiments

DTD (5760) — Robustness & Scalability



Method	Quality ↓		Diversity ↑
	Floral	DTD	Floral
Gatys *et al.* [5]	90.1	N/A	433.2
MultiTexture [13]	719.8	6639.0	385.9
WCT [14]	2042.5	6673.2	309.3
PO (ours)	645.9	4672.5	397.1

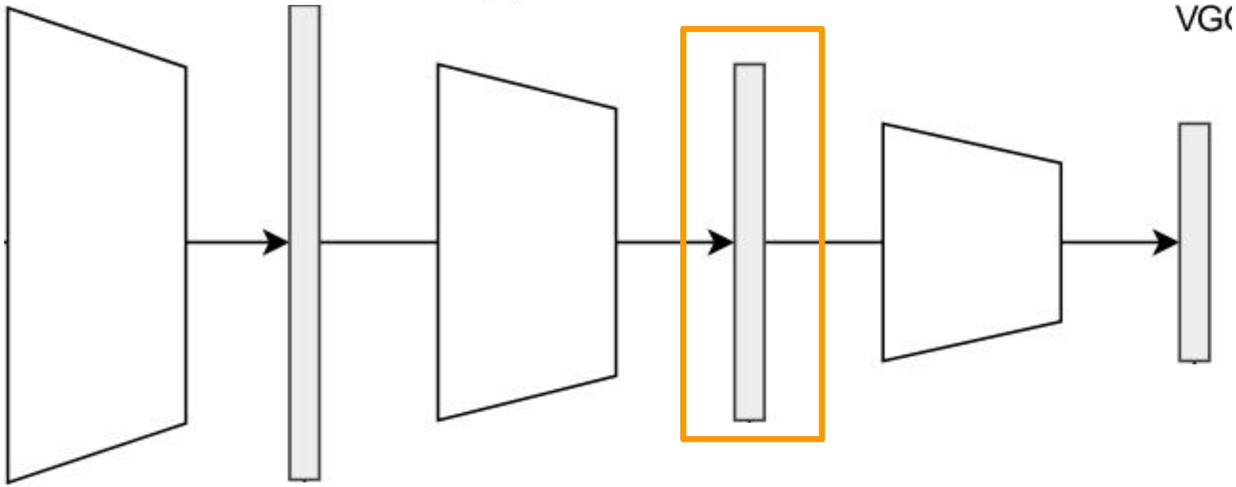
Methods comparison

Diversity

Diversity = Diversity Loss

(Single-Layer Feature Map)

$$\mathcal{L}_{div}(\{x_i\}_{i=1}^{b}) = \mathop{\mathbb{E}}_{i \neq j} \left\| F^{(div)}(x_i) - F^{(div)}(x_j) \right\|_1$$



Low level features

Abstract features

High abstract Task specific Features

Ablation study

AdaPO vs ProPO

AdaPO: Frame-like artifacts

ProPO: Gradual matching

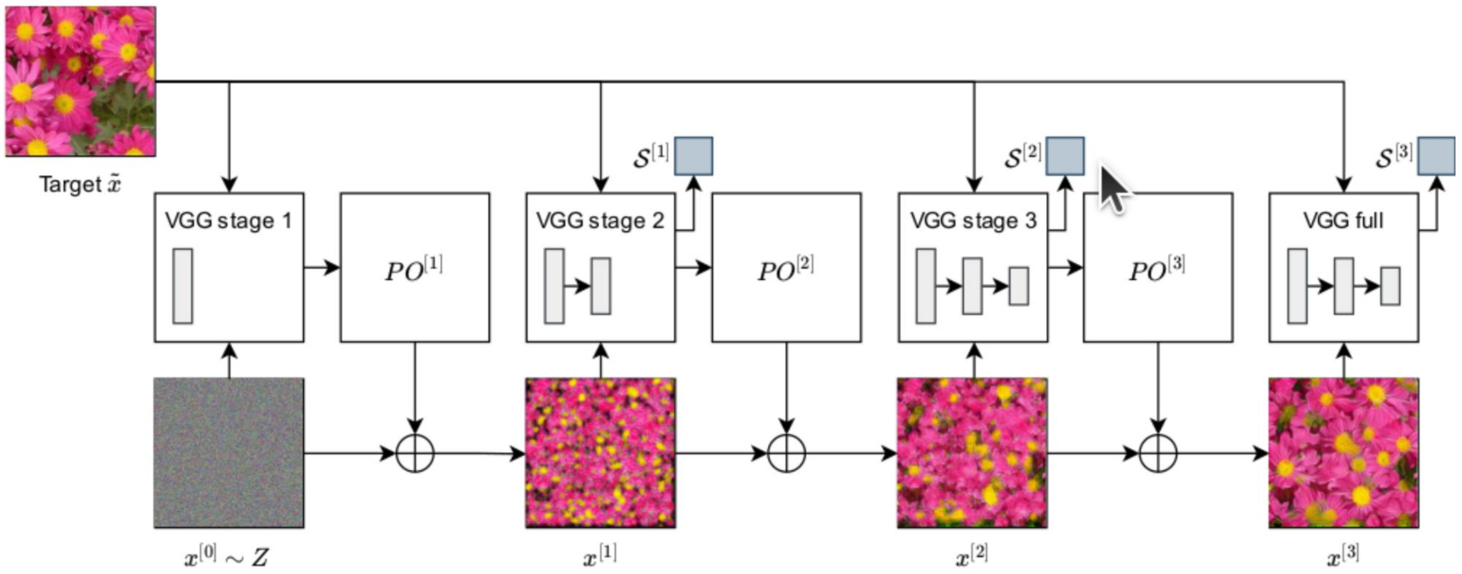
Low level features
(i.e. colors)

Little modifications
(richer diversity)

Ablation study

ProPO

ProPO: Gradual matching



Low level features (i.e. colors)

Little modifications (richer diversity)

ProPO: Strong points

Texture and colors

ProPO | Target | Gastys et al.

ProPO | Target | Gastys et al.

ProPO (Sample 1) | ProPO (Sample 2) | ProPO (Sample 3) | Target

ProPO: Limitations

Large-scale patterns

ProPO Target Gastys et al.

ProPO Target Gastys et al.

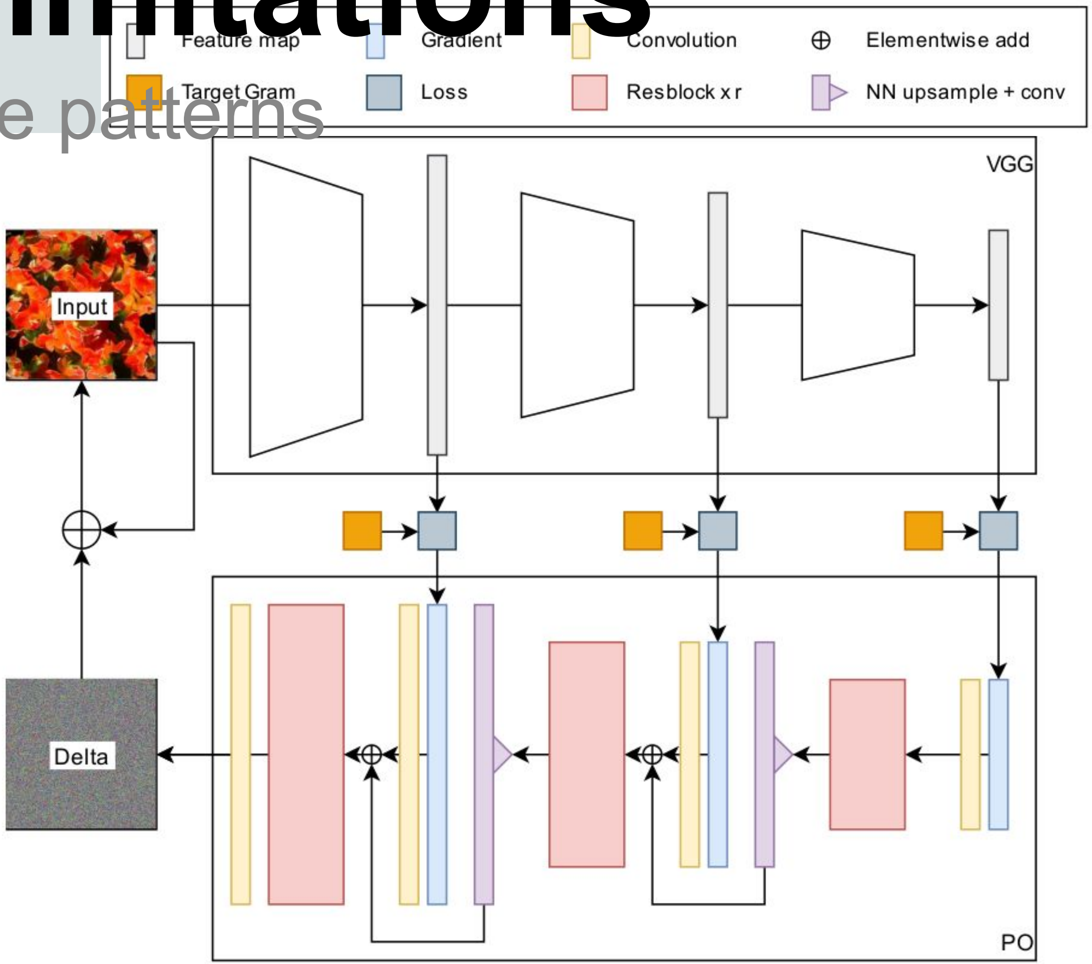
ProPO Target Gastys et al.

ProPO (Sample 1) ProPO (Sample 2) ProPO (Sample 3) Target

ProPO: Limitations

Large-scale patterns

Method output Target



Why do you think the algorithm fails to preserve some structural components of the target image?

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

Thanks for your attention!
Any questions?

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

Don’t worry, we picked them for you!

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

Jérôme SIOC'HAN DE KERSABIEC:

Why did they use 5 stages in the model architecture and not 4 or 6. What are the effects of choosing a specific number of stages on results ?

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

Maissa Beji:

How does the design of the Pseudo Optimizer (PO) network address the challenge of reducing computation time while maintaining quality in texture synthesis, as described in section3.2?

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

Santiago QUINTEROS:

If you want to find the texture synthesis of an image on the internet, what is the input you give to the network? Can you use an arbitrary image?

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