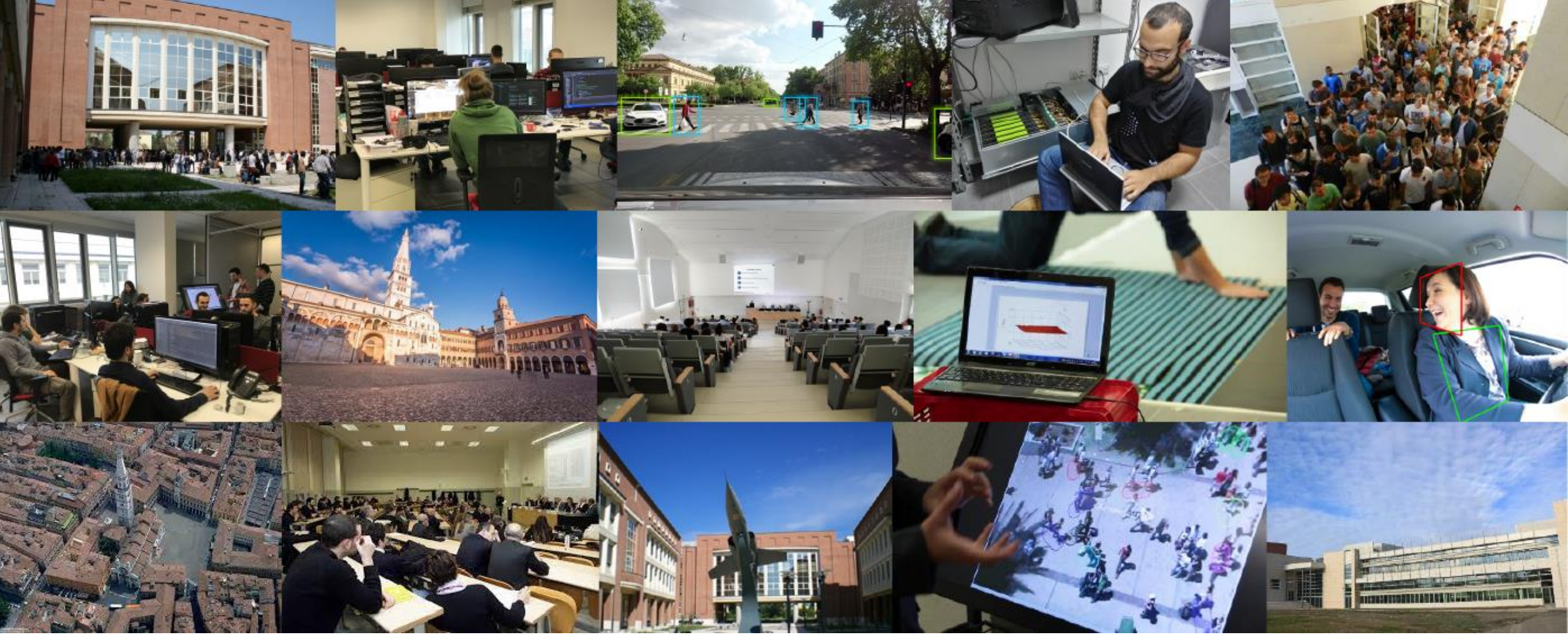


# What was Monet seeing while painting? Translating artworks to photo-realistic images

M. Tomei, L. Baraldi, M. Cornia, R. Cucchiara

AImage Lab

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

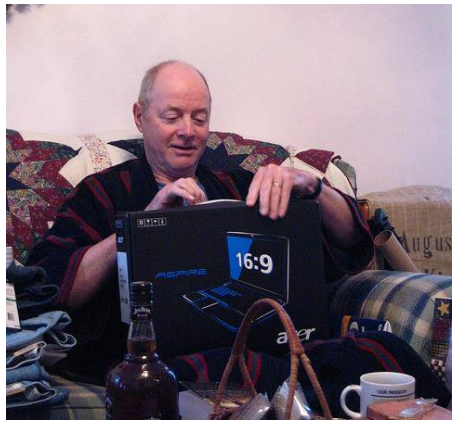
# COMPUTER VISION IN THE ARTISTIC DOMAIN

AImage Lab

The effectiveness of Computer Vision solutions has greatly increased in the last years, also thanks to Deep Learning architectures.

**However, if we try to apply them to the artistic domain:**

- Much of the development of recent years is due to datasets with **natural images**
- Biases in the trained models → limited applicability to the **artistic** domain ☹
- Lack of large annotated datasets to overcome the domain gap ☹




Domain shift

# HOW DOES A CNN SEE ART?

AImage Lab

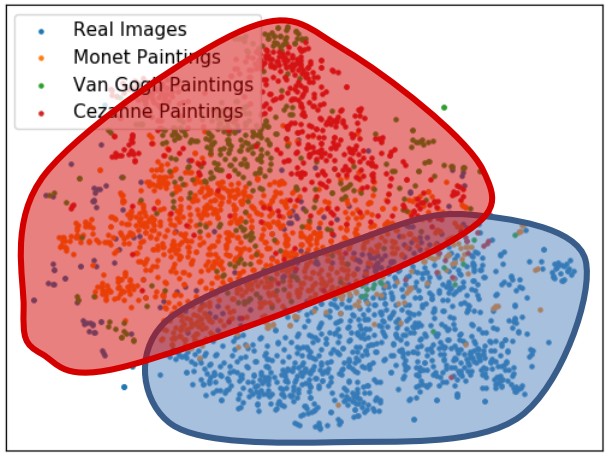
*Meadow with Poplars*
Monet, 1875

Real landscape

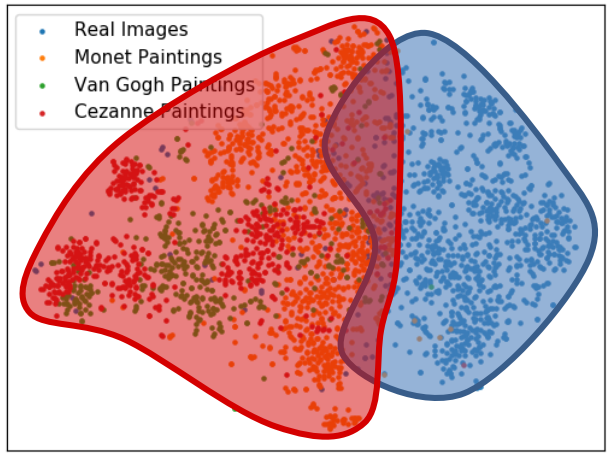
Pre-trained
CNN

Trained on real data!

Real landscapes
Paintings with landscapes




VGG-19 activations




ResNet-152 activations

Same semantic content
**Domain shift between artistic and real data**
(and also between different artists)

# METHOD OVERVIEW

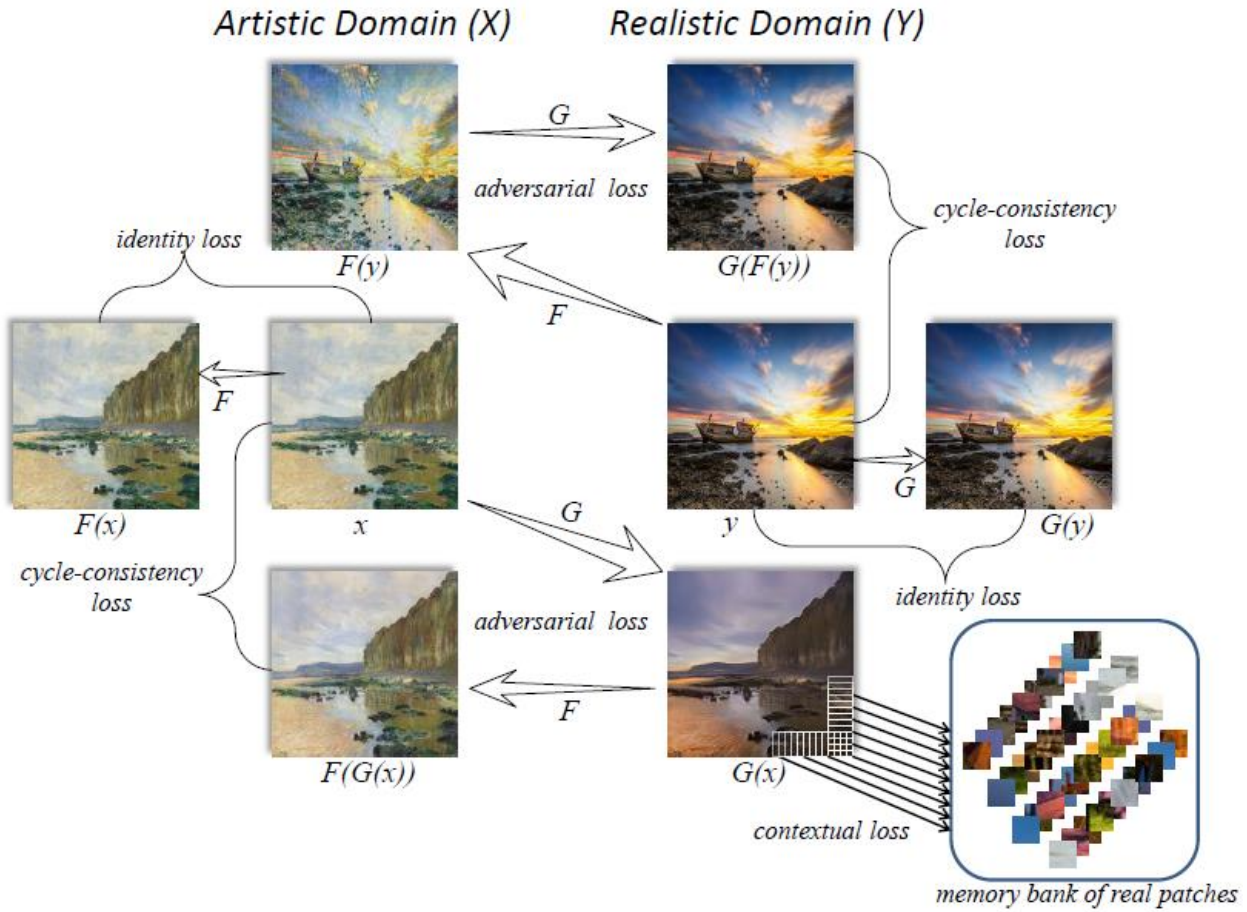
We reduce the domain gap between artistic and real data at the pixel level, by transforming artistic images into more realistic ones.

→ An unpaired domain translation task!

**Strategy**:

→ We build upon the **Cycle-GAN** framework for domain translation

→ … and augment it with a patch-level retrieval strategy which lets us **"copy" from real images**.



# THE CYCLE-GAN FRAMEWORK

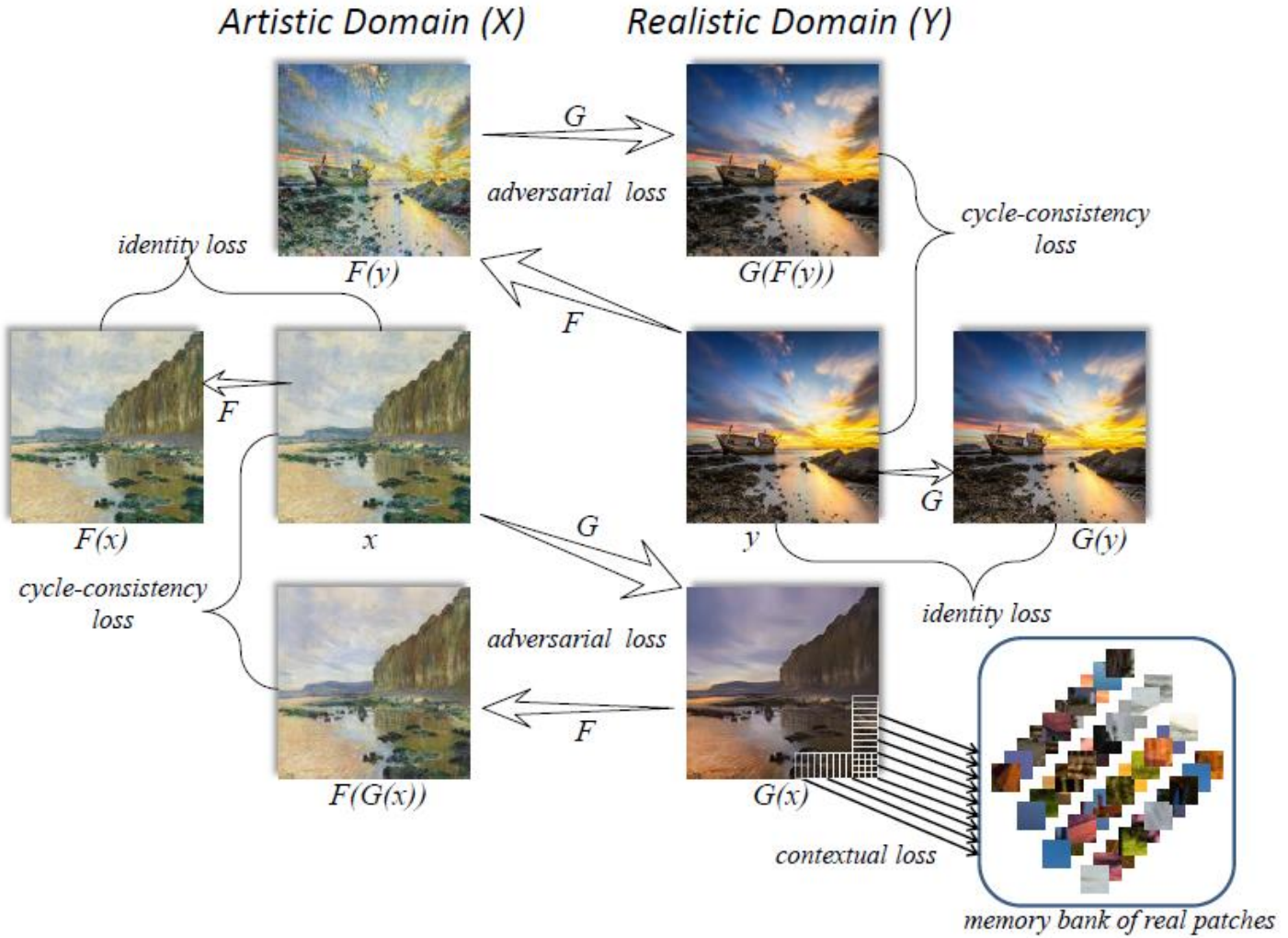
- **Two adversarial losses**: in both directions, generators are trained to reproduce the target data distribution, and the discriminator to distinguish between real and generated images

- Additional constraint: **cycle consistent loss**

$$x \rightarrow G(x) \rightarrow F(G(x)) \approx x$$

$$y \rightarrow F(y) \rightarrow G(F(y)) \approx y$$

- **Identity loss**: to preserve the colour distribution, forces the generator to behave like an identity function when given images from the target distribution



# COPYING FROM THE REAL

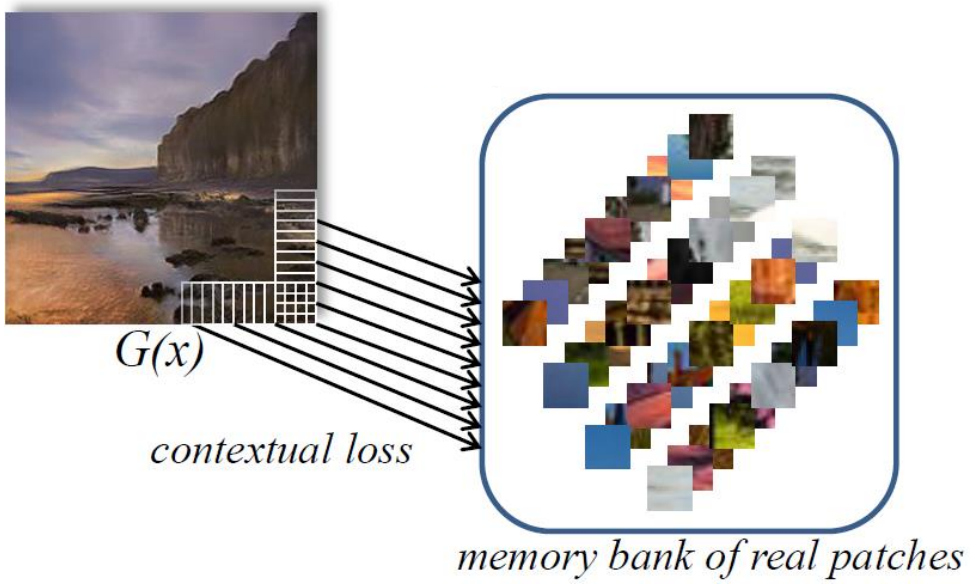
AImage Lab

Cycle-GAN discriminators and generators do a great job but…

How do we ensure that every detail of generated images look real?

→ **We can copy from real images**!

1) Build a **memory bank of real patches**
2) Split *G*(*x*) into patches as well
3) Pair each patch in *G*(*x*) with its most similar real patch, and **maximize their similarity**

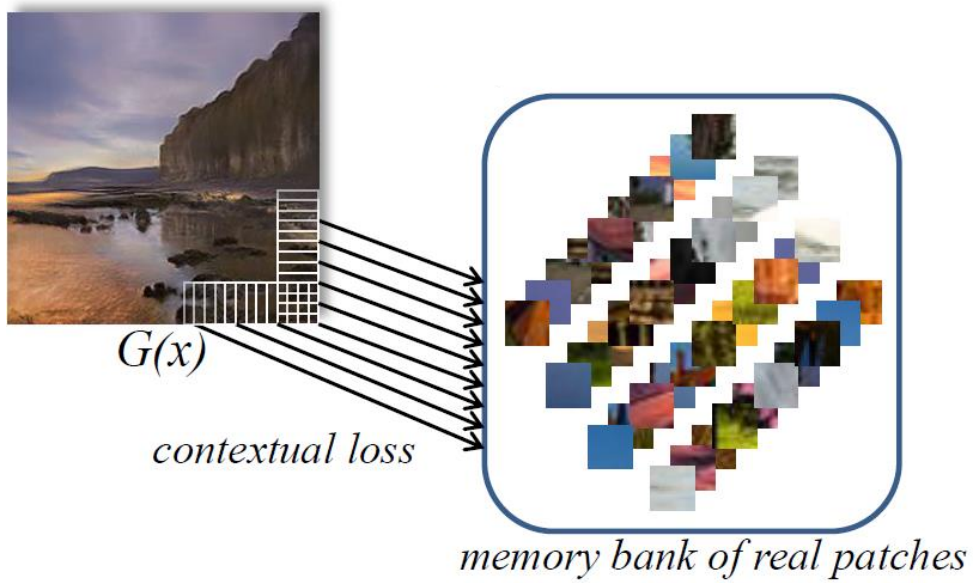


# COPYING FROM THE REAL

AImage Lab

1) Build a **memory bank of real patches**
2) Split $G(x)$ into patches as well
3) Pair each patch in $G(x)$ with its most similar real patch, and **maximize their similarity**

→ we have an additional loss which needs to get evaluated at each iteration!




Issues…

- How do we maximize the similarity
- How do we efficiently retrieve the most similar patch

# RETRIEVING REAL PATCHES

Define a **pairwise distance function** between real and generated patches

$$d_{ij} = \left(1 - \frac{(k_i - \mu_m) \cdot (m_j - \mu_m)}{\|k_i - \mu_m\|_2 \, \|m_j - \mu_m\|_2}\right), \text{ where } \mu_m = \frac{1}{N} \sum_j m_j$$

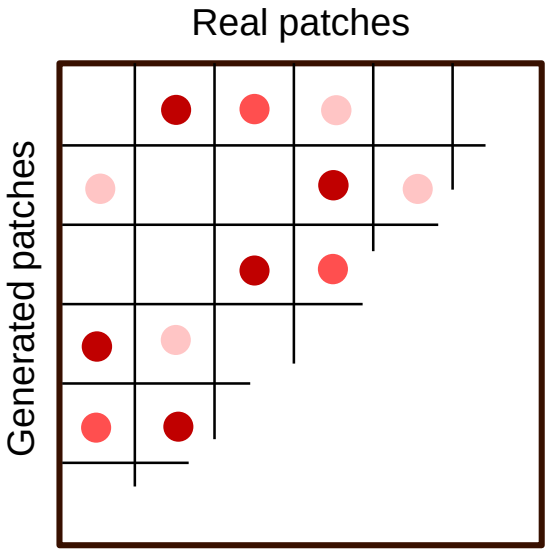
Generated patches

Real patches

Then, do a **(soft) assignment!**

$$\tilde{d}_{ij} = \frac{d_{ij}}{\min_l d_{il} + \epsilon}, \text{ where } \epsilon = 1e-5$$

$$A_{ij} = \frac{\exp(1 - \tilde{d}_{ij}/h)}{\sum_l \exp(1 - \tilde{d}_{il}/h)} = \begin{cases} \approx 1 & \text{if } \tilde{d}_{ij} \ll \tilde{d}_{il} \; \forall \, l \neq j \\ \approx 0 & \text{otherwise} \end{cases}$$




→ Each generated patch is assigned prominently to its mostly similar real patch, and to others which have high affinity degrees

# REDUCING THE COMPUTATIONAL OVERHEAD

The size of the pairwise affinity matrix grows linearly with the number of patches

→ **computationally intractable**

## Solution

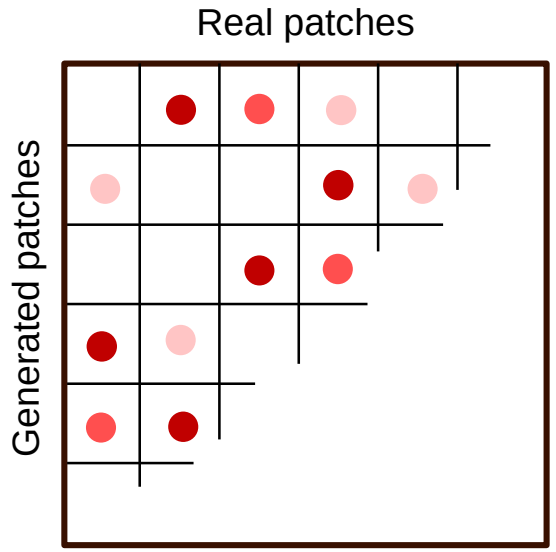
Build a **suboptimal Nearest Neighbour index** with real patches

For each generated patch, retrieve the *k* most similar

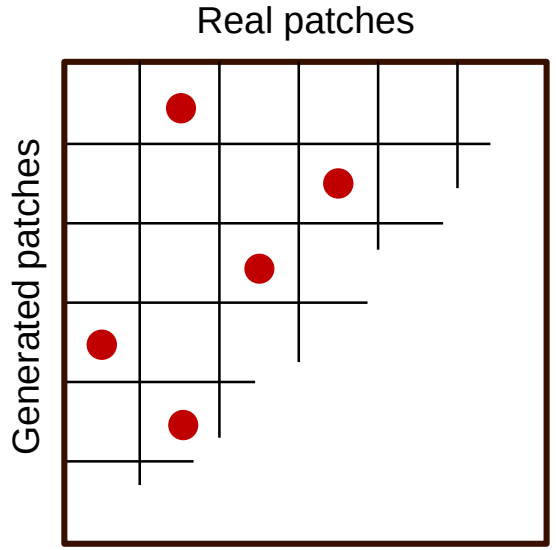
Estimate the affinity matrix using the retrieved real patches

→ A sparse matrix with 0 on non-retrieved patches

→ Very similar to the true matrix ☺




Exact matrix




Approx. matrix ($k$ = 1)

# MAXIMIZING THE SIMILARITY WITH REAL PATCHES

AImage Lab

We maximize the similarity between each generated patch and its corresponding assignment from the memory bank, using the **contextual loss** [1]:

$$\mathcal{L}_{CX}(K, M) = -\log\left(\frac{1}{N}\left(\sum_i \max_j A_{ij}\right)\right)$$

## Multi-scale version

We build multiple memory banks for multiple patch scales

→ The final objective is the sum of the losses obtained at each scale

[1] Mechrez, R., Talmi, I., Shama, F., Zelnik-Manor, L.: Learning to maintain natural image statistics. arXiv preprint arXiv:1803.04626 (2018)

# EVALUATION

AImage Lab

**Fréchet Inception Distance** (FID)

→ Wasserstein-2 distance between two Gaussians fitted on real and generated data, using Inception-v3 activations

→ more consistent to human judgment than the Inception Score

→ *Higher the FID, higher the realism of the generation* (according to a CNN ☺)

We compare with:

(a) The style transfer of real images using the style of the paintings
(b) Cycle-GAN

Using both low-level and higher-level activations from Inception-v3.

→ Increased realism at all levels and with all artists.

| | **Monet** | **Cezanne** | **Van Gogh** |
|---|---|---|---|
| 2048 dimensions | | | |
| Original paintings | 74.45 | 176.51 | 166.72 |
| Style Transfer [1] | 58.02 | 91.23 | 101.54 |
| CycleGAN [2] | 55.26 | 83.62 | 86.82 |
| Our Model | **54.43** | **77.01** | **81.74** |
| 768 dimensions | | | |
| Original paintings | 0.52 | 1.26 | 1.39 |
| Style Transfer [1] | 0.50 | 1.01 | 1.18 |
| CycleGAN [2] | 0.41 | 0.49 | 0.48 |
| Our Model | **0.34** | **0.37** | **0.41** |
| 192 dimensions | | | |
| Original paintings | 0.94 | 1.67 | 3.96 |
| Style Transfer [1] | 0.71 | 1.49 | 3.33 |
| CycleGAN [2] | 0.31 | 0.28 | 0.19 |
| Our Model | **0.16** | **0.13** | **0.11** |

[1] Gatys, L.A., Ecker, A.S., Bethge, M.: "Image style transfer using convolutional neural networks." *CVPR* 2016.
[2] Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. ICCV 2017

# EVALUATION

AImage Lab

## User Study

- **Realism of the generation**: given corresponding images generated by CycleGAN and our approach, select the most realistic
- **Coherency with the painting**: given the painting and the results of CycleGAN and of our approach, select the most faithful to the original painting

| Test | Scale | CycleGAN [1] | Our method |
|---|---|---|---|
| Realism of the generation | $256 \times 256$ | 41.6% | 58.4% |
| Coherency with the painting | $256 \times 256$ | 41.2% | 58.8% |

(percentage of times a method has been selected)

[1] Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. ICCV 2017

# EVALUATION

AImage Lab

## User Study

- **Multi-scale comparison with real images**: given a random real image and one generated by our approach, select the one that seems more realistic

  Repeated when resizing with a ratio of 1, 2/3 and ½

| Scale | Random real image | Generated image |
|---|---|---|
| $256 \times 256$ | $95.1\%$ | $4.9\%$ |
| $170 \times 170$ | $88.2\%$ | $11.8\%$ |
| $128 \times 128$ | $88.0\%$ | $12.0\%$ |

→ sometimes we can fool the user! ☺

# QUALITATIVE RESULTS

AImage Lab

| Original painting | Cycle-GAN | Ours |
| --- | --- | --- |
|  |  |  |
|  |  |  |

# QUALITATIVE RESULTS

AImage Lab

Original painting

Cycle-GAN

Ours

# CONCLUDING

AImage Lab

The domain shift between artistic and real data limits the performance of pre-trained CNNs on the cultural domain. We propose to solve it by translating artistic images to the real domain.

**Contributions**:

- A novel method for artistic-to-realistic domain translation
- The generation is driven at patch-level using multi-scale memory banks + efficient implementation which employs approximate NN search
- Qualitative and quantitative results outperforms the Cycle-GAN baseline, leading to more realistic results

# (MORE) QUALITATIVE RESULTS

AImage Lab








# (MORE) QUALITATIVE RESULTS

AImage Lab








UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Centro Interdipartimentale di Ricerca
Softech: ICT per le Imprese

AImage Lab

# Thank you!

marcella.cornia@unimore.it
aimagelab.ing.unimore.it


Matteo Tomei


Lorenzo Baraldi


Marcella Cornia


Rita Cucchiara

Facebook AI Research has selected AImageLab as one of the 15 world-class research labs in Europe

Thanks to the **“CultMedia” Project** of the National Technology Cluster on Smart Communities, cofounded by the Italian Ministry of Education, University and Research (MIUR).