Generative Models for Image-to-Image Translation and Continual Learning

Gabriele Graffieti
gabriele.graffieti@unibo.it

PhD Student
Computer Science and Engineering Department
University of Bologna
Italy

July 9, 2020
Generative models

## In a nutshell

- They model the join probability $P(X, Y)$.
- They learn the distribution of the data.
- More general than discriminative models (a DM can always be derived from a GM).
- Since they learn a distribution as similar as possible to the real distribution of data, it is possible to sample artificial data from them.
- They can handle multi-modal output, where more than one $Y$ is the correct prediction for a single $X$ (e.g. $X$ is a frame in a video, and $Y$ is the predicted next frame).
- Usually unsupervised.
Generative models
Image-to-Image translation

• Like a language translation, we want to translate one image from one domain to another, maintaining unchanged the semantic content.
• The ground truth result of the translation may not be known, and more than one result might be correct.
Most used models for image-to-image translation

- If you have ground truth data (paired images) → pix2pix [1]
- If you have unpaired data → cycleGAN [2]
  - The problem is similar to the problem of learning to translate between 2 languages without a dictionary.
  - We could exploit the *cycle consistency* properties of mappings:

\[
F(G(x)) \approx x \quad \text{and} \quad G(F(y)) \approx y
\]

- We force the model to learn the *inverse mapping*, and the composition of the 2 translation have to be similar to the input.
Defogging
Defogging

- Applicable to many fields (self-driving cars, automotive, surveillance, aesthetic quality of photos...).
- Intrinsically unpaired:
  - Almost impossible to have the same scene with and without fog.
  - Adding artificial fog requires exact scene depth (difficult to obtain in an outdoor scenario).
  - Using man-made fog is onerous, time consuming and cannot be done in many scenarios (e.g. roads).
- We are interested in defogging images with thick and severe fog (many works and datasets concentrate on light haze).
Defogging thick fog using paired/unpaired data

(a) Real

(b) Artificial fog

(c) Totally unpaired
CurL-Defog [3]

Paired Images

\[ \mathcal{L}_{CL} \]

\[
\begin{align*}
G_{defog} & \quad \rightarrow \quad \hat{c} \\
D_{clear} & \quad \rightarrow \quad c
\end{align*}
\]

Paired Images

\[
\begin{align*}
G_{fog} & \quad \rightarrow \quad f \\
D_{fog} & \quad \rightarrow \quad c
\end{align*}
\]

\[ \Phi(f) \] \quad Cycle Perceptual Loss (\( \mathcal{L}_{perc} \))

\[ \Phi(f_{rec}) \] \quad Cycle Consistency Loss (\( \mathcal{L}_{cy} \))

[Diagram showing VGG-16 and other components related to cycle consistency and perceptual loss.]
CurL-Defog [3]

(a) Real image
(b) Artificial fog
(c) Totally unpaired
(d) CurL-Defog
Correction of face morphing artifacts
Printed & scanned morphed images [4]

- Demorphing is usually performed on high-resolution digital morphed images.
- But when we request a document we have to provide a small printed photo ID that is scanned electronically.
- Printing and scanning images degrade their quality, and some morphing artifacts may be not visible after the process.
Simulation of printed & scanned morphed images

- Training DL models on printed & scanned morphed images require a large dataset.
- Printing and scanning a photograph is time consuming and cannot be completely automatized.
- We want to exploit image-to-image translation techniques to performing the mapping automatically, without the need to manually perform the printing and the scanning of the images.
Straightforward technique: just store past data and repeat them through the network.

.Requires extra storage (e.g. for ImageNet, if we store 20 patterns per class, the total storage is about 3.8 GB)

Requires extra forward/backward steps when mixing new and old patterns more iterations for epoch.

Not really biologically plausible.
Continual learning with latent replay [5]

- **Efficiency**: extra forward and backward steps take place only in the upper layers.
- **More biologically plausible**.
- **Less storage required**.
- **Features aging**.

---

Gabriele Graffieti

Generative models for img2img translation and CL

July 9, 2020 15 / 18
Latent generative replay

- Class specific discriminative features (training at full pace)
- Low-level generic features (slow training)

Output Layer (classes)
Latent replay layer
Generative Model
Input Layer (images)

Efficiency: extra forward and backward steps take place only in the upper layers.
No storage required.
No features aging.
No constraints on the number of replay patterns.
Even more biologically plausible.
Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros
Image-to-Image Translation with Conditional Adversarial Networks
CVPR 2017

Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks
ICCV 2017

Gabriele Graffieti, Davide Maltoni
Towards Artifacts-free Image Defogging
To appear, ICPR 2020
Matteo Ferrara, Annalisa Franco, Davide Maltoni
Face morphing detection in the presence of printing/scanning and heterogeneous image sources
arXiv Preprint, 2019

Lorenzo Pellegrini, Gabriele Graffieti, Vincenzo Lomonaco, Davide Maltoni
Latent Replay for Real-Time Continual Learning
To appear, IROS 2020