CycleGAN

ISPR - Midterm 4

Filippo Baroni

September 2021

Task

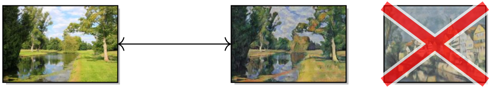
Unsupervised domain translation

Given samples $\{$, , $\ldots\} \subseteq \mathcal{X}$ and $\{$, , $\ldots\} \subseteq \mathcal{Y}$ from two different domains, learn a mapping

$$G : \mathcal{X} \longrightarrow \mathcal{Y}$$

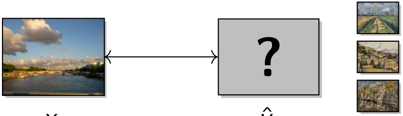
such that $\hat{y} := G(x)$ is indistinguishable from $y \in \mathcal{Y}$.

- **Conditional generation:** the output $\hat{y} \in \mathcal{Y}$ should retain some features of the input $x \in \mathcal{X}$.



x $\hat{y}$

- **Unpaired samples:** the expected output $\hat{y}_i$ for a specific sample $x_i \in \mathcal{X}$ is unknown even at training time.



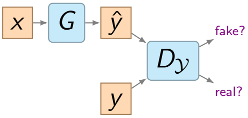
x_i $\hat{y}_i$

Loss function

The innovation of cycle-consistency

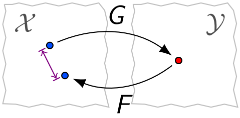
Idea: train two generators $G\colon \mathcal{X} \to \mathcal{Y}$ and $F\colon \mathcal{Y} \to \mathcal{X}$ simultaneously.

GAN architecture



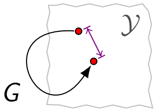
$G(x) \approx y$

Cycle-consistency



$F \circ G = \mathrm{id}_{\mathcal{X}}$

Identity regularization



$G(y) = y$

$$
\begin{aligned}
\mathcal{L} &= \mathbb{E}_{x\sim p(x)}[\log(1 - D_{\mathcal{Y}}(G(x)))] + \mathbb{E}_{y\sim p(y)}[\log D_{\mathcal{Y}}(y)] && \mathcal{L}_{\text{adv}}(G, D_{\mathcal{Y}}) \\
&\quad + \mathbb{E}_{y\sim p(y)}[\log(1 - D_{\mathcal{X}}(F(y)))] + \mathbb{E}_{x\sim p(x)}[\log D_{\mathcal{X}}(x)] && \mathcal{L}_{\text{adv}}(F, D_{\mathcal{X}}) \\
&\quad + \lambda \cdot \mathbb{E}_{x\sim p(x)}[\|F(G(x)) - x\|_1] && \mathcal{L}_{\text{cyc}}(F \circ G) \\
&\quad + \lambda \cdot \mathbb{E}_{y\sim p(y)}[\|G(F(y)) - y\|_1] && \mathcal{L}_{\text{cyc}}(G \circ F) \\
&\quad + \alpha \cdot \mathbb{E}_{y\sim p(y)}[\|G(y) - y\|_1] && \mathcal{L}_{\text{id}}(G) \\
&\quad + \alpha \cdot \mathbb{E}_{x\sim p(x)}[\|F(x) - x\|_1] && \mathcal{L}_{\text{id}}(F)
\end{aligned}
$$

Model

Architectures of the generators and the discriminators

Results

Conclusion

Strengths and weaknesses

Pros

- Multi-purpose image-to-image domain translation.
- Fully unsupervised model.

Cons

- Failure to adapt to unseen contexts.
- Only suitable for texture/style changes.