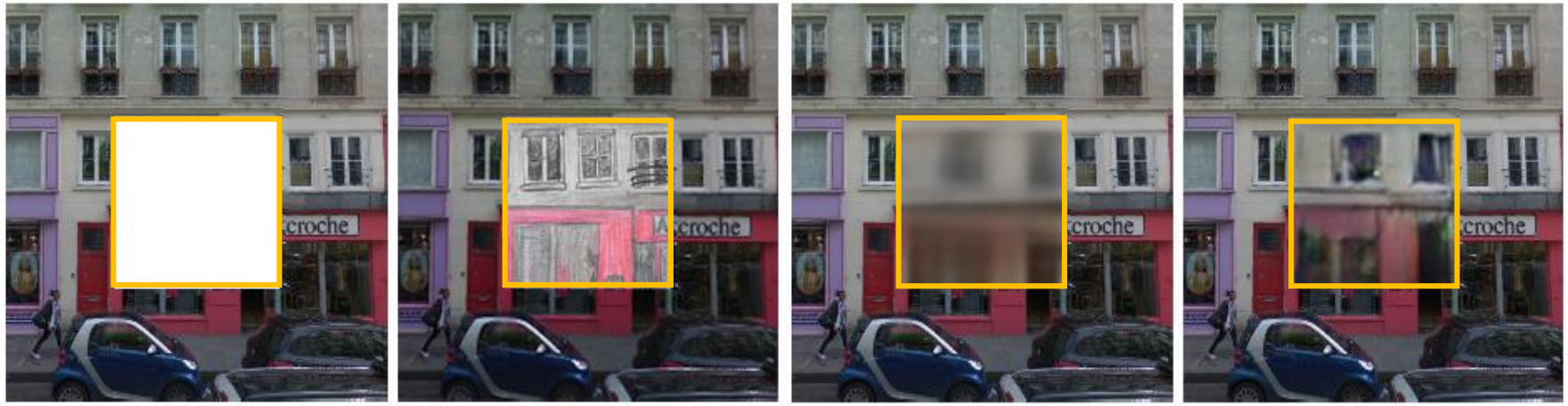
Context Encoders: Feature Learning by Inpainting

7107053114 沈佳詠

Abstract

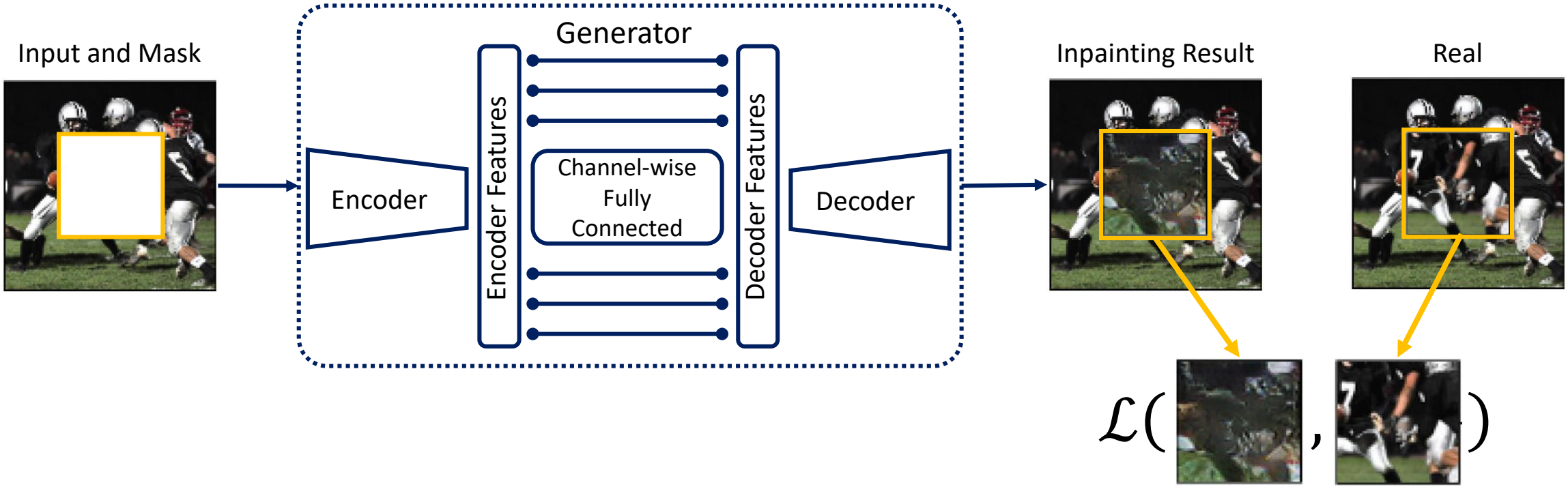


(a) Input context (b) Human artist (c) Context Encoder ($L2$ loss) (d) Context Encoder ($L2$ + Adversarial loss)

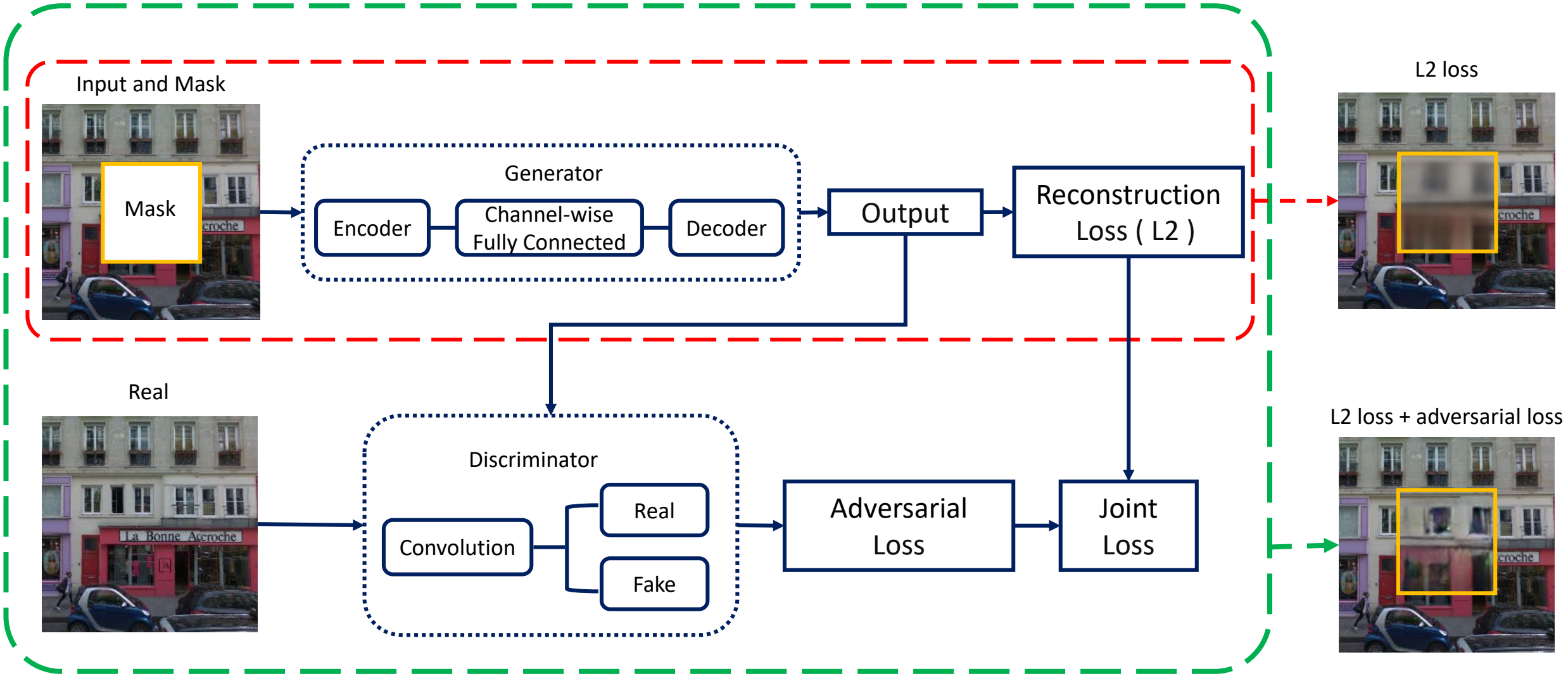
- **Approach**：
  - An unsupervised visual feature learning algorithm driven by context-based pixel prediction.
- **Models**：
  - Convolutional Neural Network
  - Conditional Generative Adversarial Network

Context Encoder

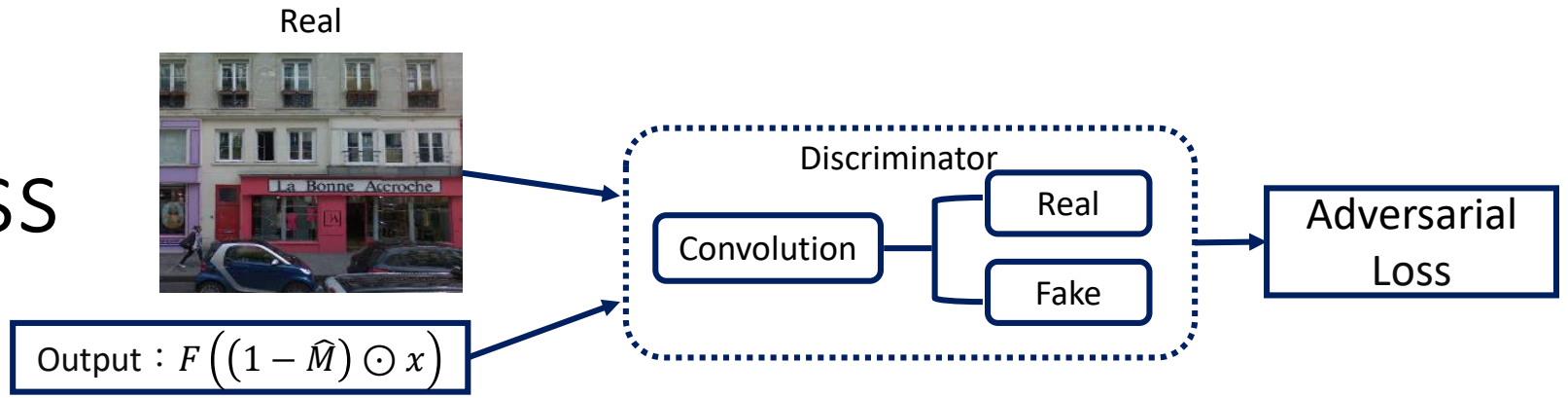
- CNNs that predict missing parts of a scene from their surroundings.



Framework


Input and Mask
Mask
Generator
Encoder
Channel-wise Fully Connected
Decoder
Output
Reconstruction Loss (L2)
L2 loss
Real
Discriminator
Convolution
Real
Fake
Adversarial Loss
Joint Loss
L2 loss + adversarial loss

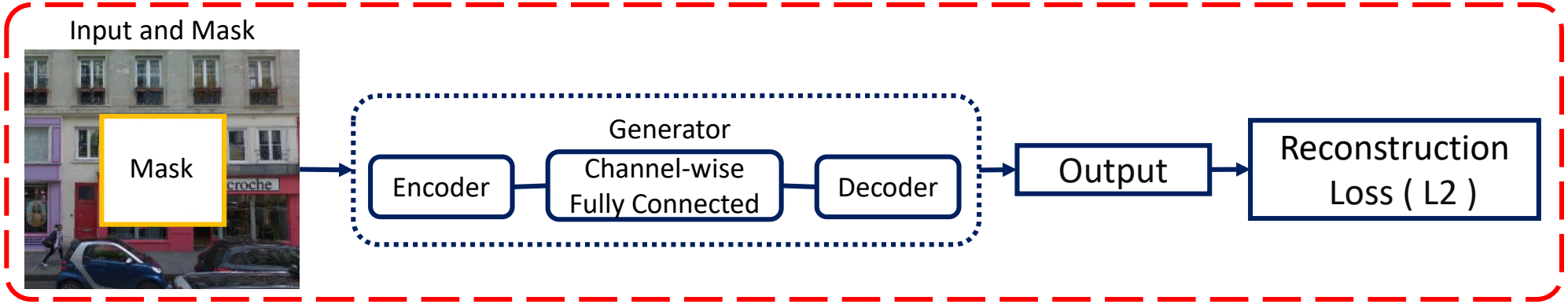
Adversarial Loss



- Original Adversarial Loss
  - $\min_G \max_D \mathbb{E}_{x\in\mathcal{X}}\left[\log\left(D(x)\right)\right] + \mathbb{E}_{z\in\mathcal{Z}}\left[\log\left(1-D\left(G(z)\right)\right)\right]$
    - G：Generator
    - x：real image
    - $\mathcal{X}$：data distribution
    - D：Discriminator
    - z：noise
    - $\mathcal{Z}$：noise distribution
- Alternate Adversarial Loss
  - $\mathcal{L}_{adv} = \max_{\mathrm{D}} \mathbb{E}_{x\in\mathcal{X}}\left[\log\left(D(x)\right) + \log\left(1-D\left(F\left(\left(1-\widehat{M}\right)\odot x\right)\right)\right)\right]$

    Generator output
    - D：Discriminator
    - F：$G \triangleq F$
    - $\widehat{M}$：input and mask，$\begin{cases} 0, & \text{for missing pixels} \\ 1, & \text{for elsewhere} \end{cases}$

Generator Loss



- Generator Loss

  Reconstruction Adversarial

  - $\mathcal{L} = \lambda_{rec}\mathcal{L}_{rec} + \lambda_{adv}\mathcal{L}_{adv}$

    - λ_{rec} 、λ_{adv} ：weights

- Reconstruction Loss (L2)

  input

  - $\mathcal{L}_{rec}(x) = \left\| \widehat{M} \odot \left(x - F\left(\left(1 - \widehat{M}\right) \odot x \right) \right) \right\|_2^2$

    - $\widehat{M}$ ：input and mask ，$\begin{cases} 0, & \text{for missing pixels} \\ 1, & \text{for elsewhere} \end{cases}$
    - x ：real image
    - F ：Generator

Expected Results

- Expected Results

Image | L2 Loss | Adversarial Loss | L2 Loss + Adversarial Loss