Image-to-Image Translation: A Literature Review

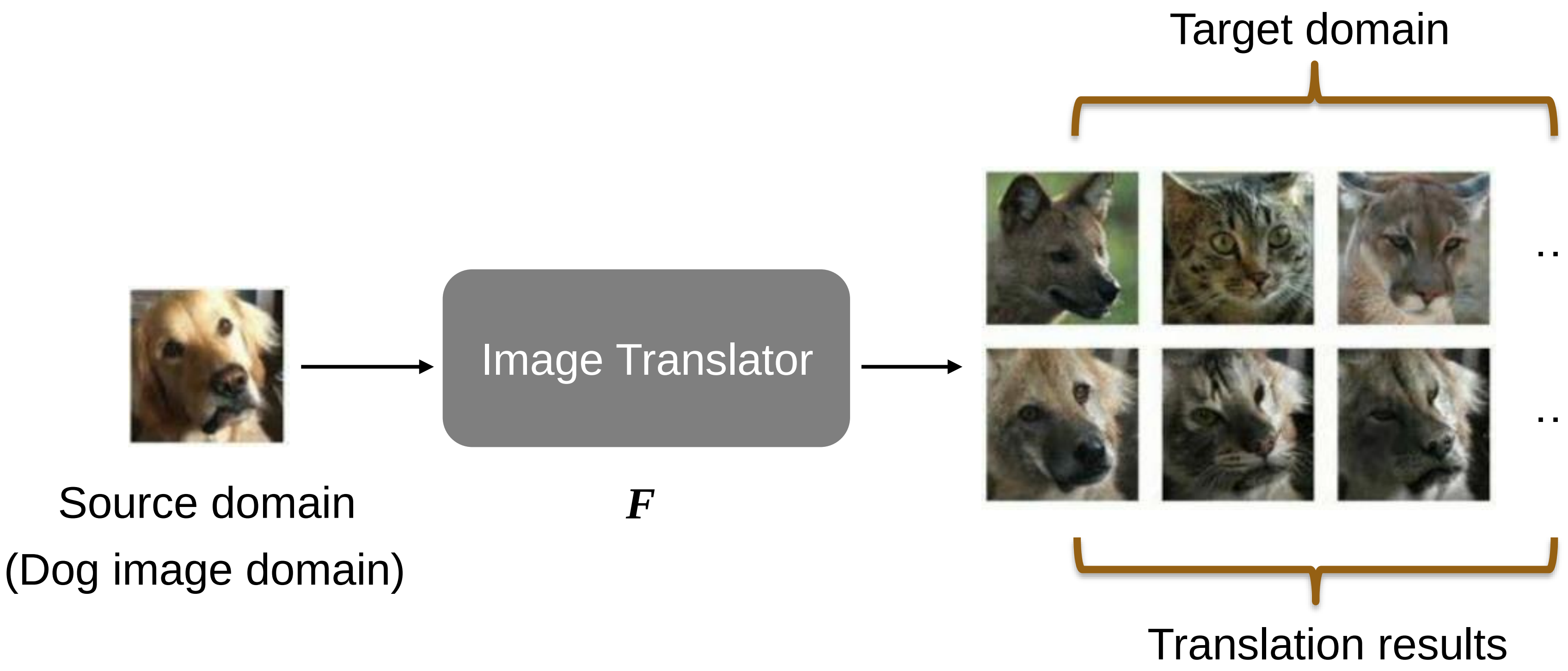
Speaker: Yuanwen Yue
Supervisor: Priyanka Chaudhary

Outline

- Introduction
- Supervised image-to-image translation
- Unsupervised image-to-image translation
- Image-to-image translation in Geomatics
- Summary

What is image-to-image translation?

- Image-to-image (I2I) translation refers to the task of mapping an image from a source domain to a target domain.



Application of I2I translation

- Sketch to photo, Blurry to sharp, Day to night, Labels to street scene, Data augmentation, etc.
- Geomatics: Aerial to map, SAR to optical image, etc.


Sketch to photo

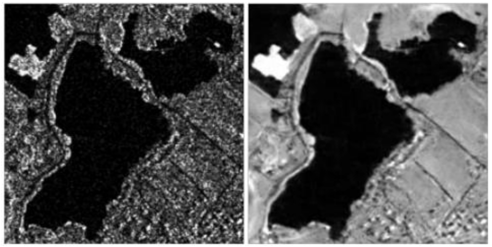

Blurry to sharp


Day to night


Gray-scale to color


Labels to street scene

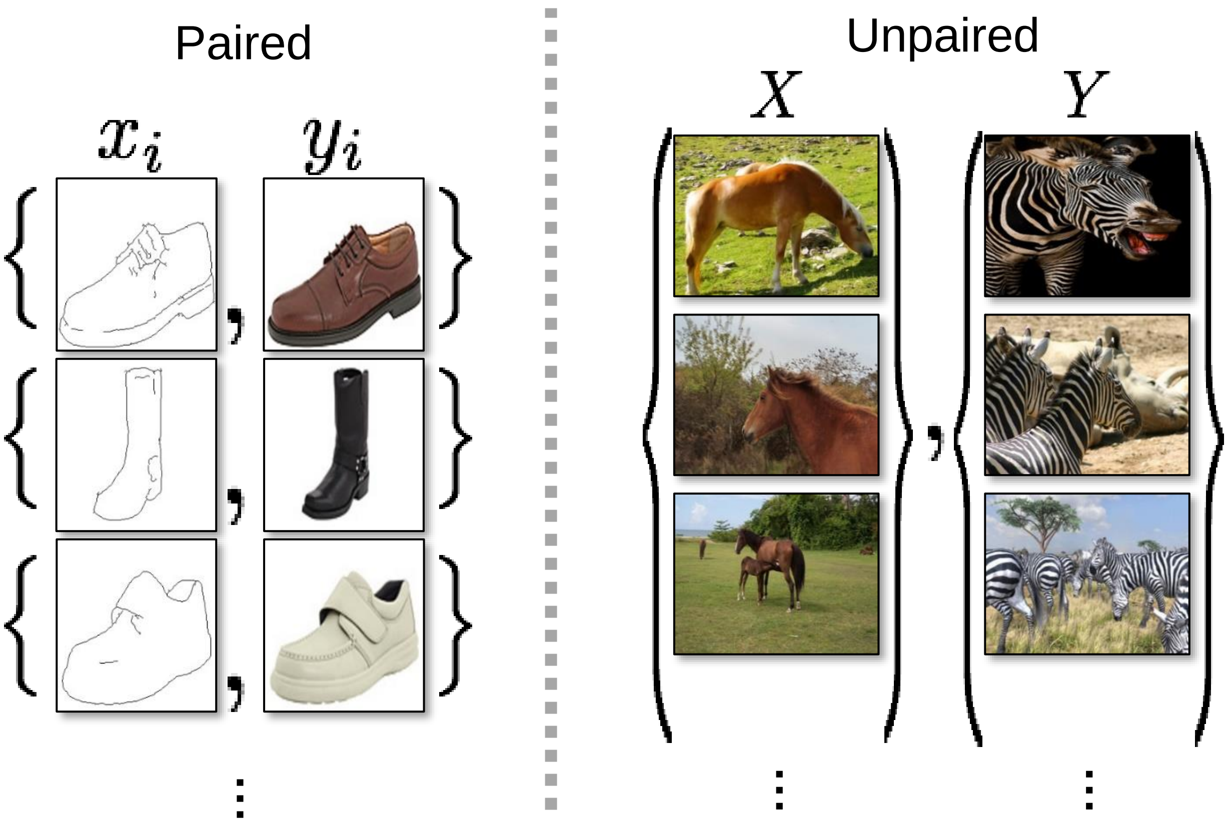

Aerial to map


SAR to optical image

...

Supervised vs. Unsupervised

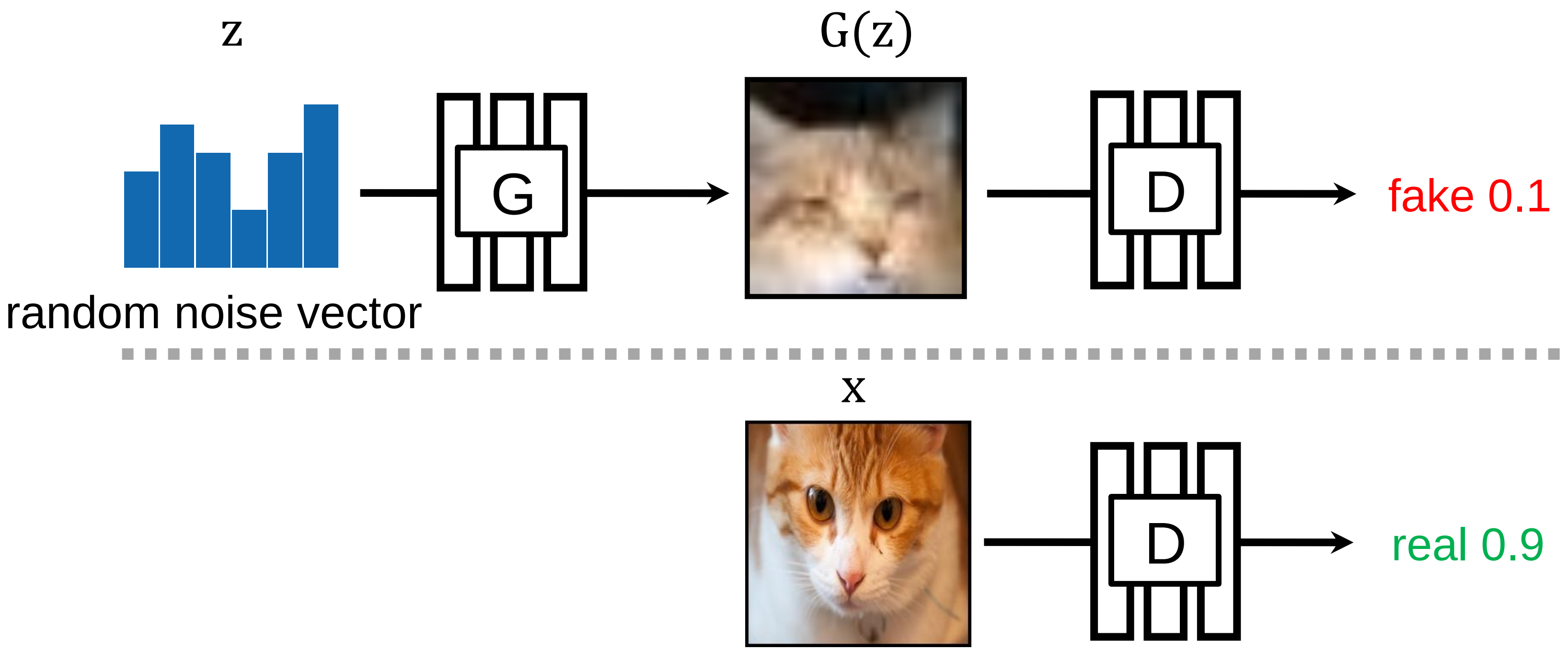
- Supervised: require paired training data as supervision.
- Unsupervised: learn the mappings between two image collections without paired training data.



[Zhu et al. 2019]

Generative Adversarial Networks (GANs)

- Generator (G): generate fake examples that can fool D.
- Discriminator (D): classify fake samples vs. real images.



$$\min_G \max_D \mathbb{E}_{z,x}[\log D(\underbrace{G(z)}_{\text{fake}}) + \log(1 - \underbrace{D(x)}_{\text{real}})]$$

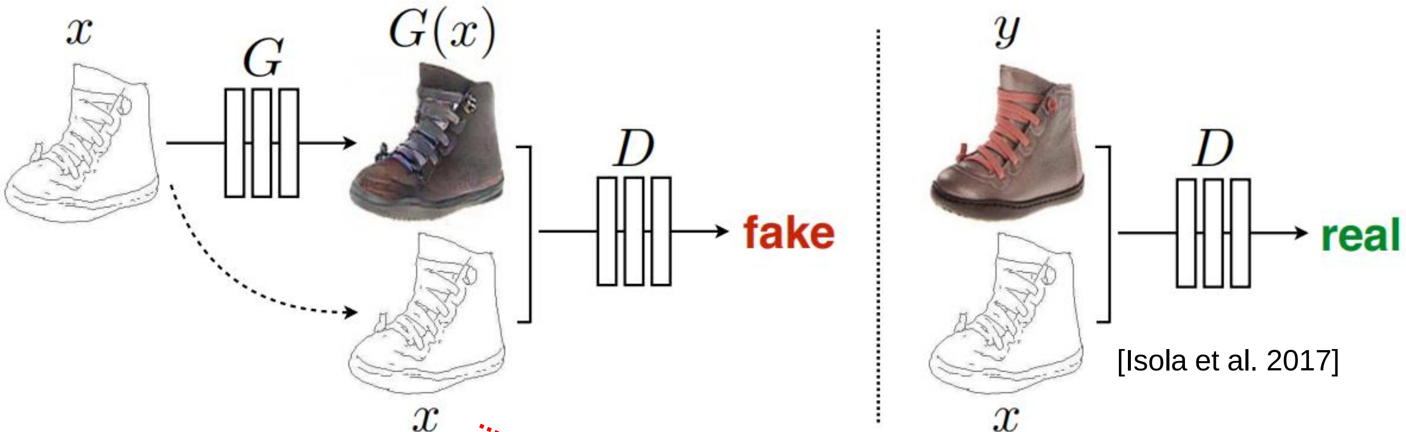
Update G

[Goodfellow et al. 2014]

Supervised Image-to-image Translation

Pix2pix: conditional GANs

- **Conditional**: condition on an input image and generate a corresponding output image.
  - Generator (G)
  - Discriminator (D)

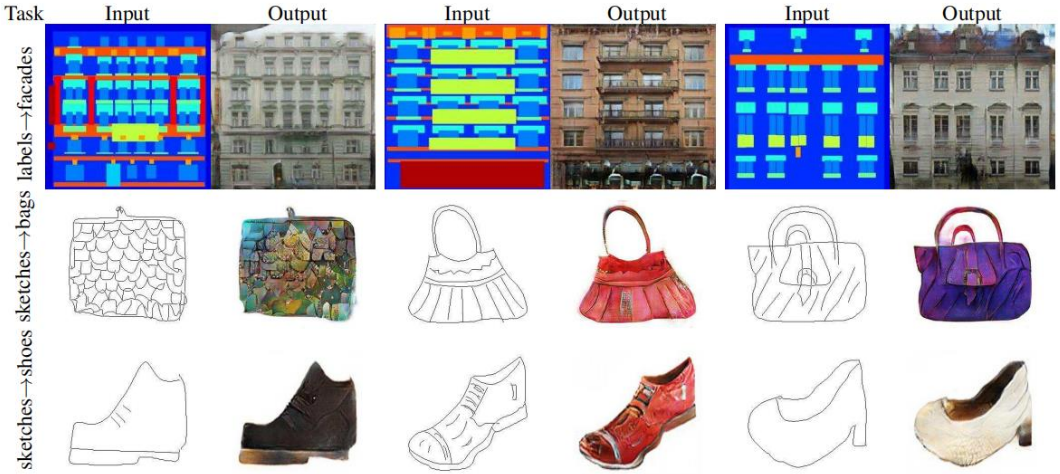


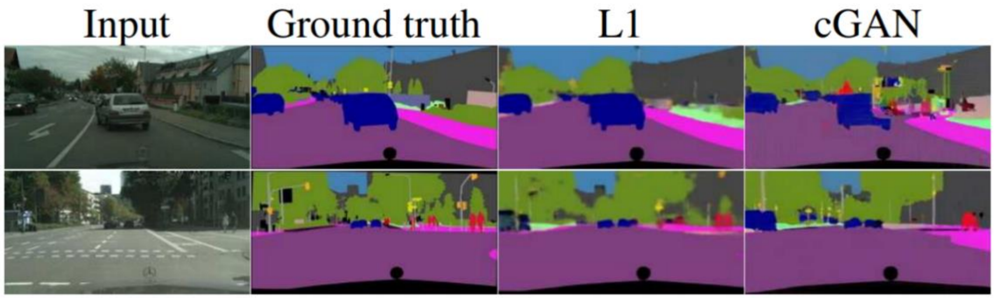
[Isola et al. 2017]

$$\min_G \max_D \mathbb{E}_{x,y}[\log \underbrace{D(x, G(x))}_{\text{fake pair}} + \log(1 - \underbrace{D(x, y)}_{\text{real pair}})]$$

Pix2pix: conditional GANs

- **Pros:**
  - Applicable in a wide variety of tasks.
  - A simple framework sufficient to achieve good results on graphics tasks, like photo generation.
- **Cons:**
  - Work not well on vision tasks, like semantic segmentation.
  - One-to-one mapping between the two domains (**Unimodality**).



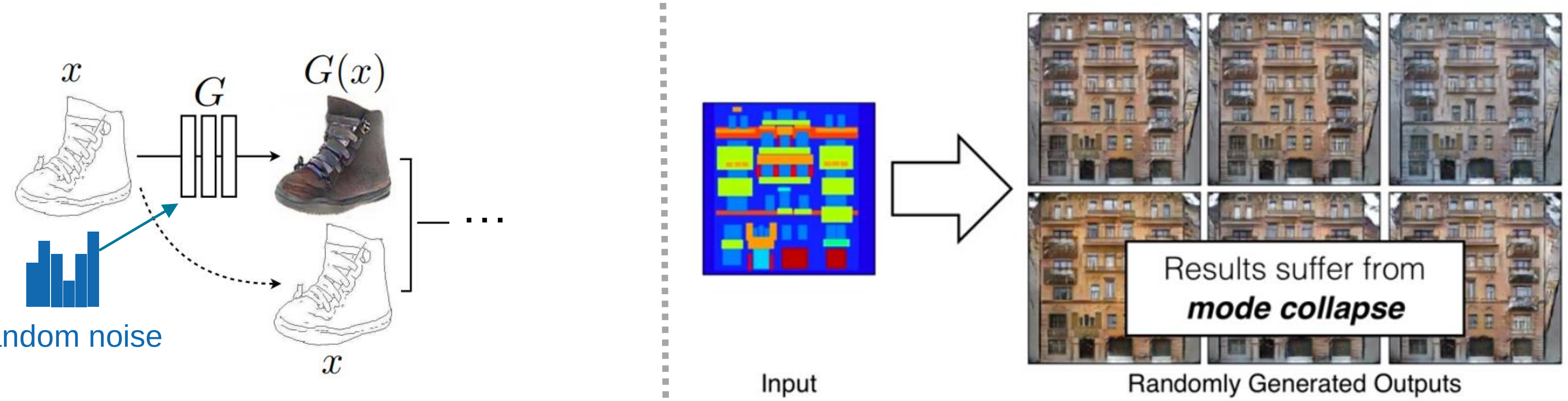


[Isola et al. 2017]

Towards Multimodal I2I Translation

- Multimodality: give multiple translation answers.

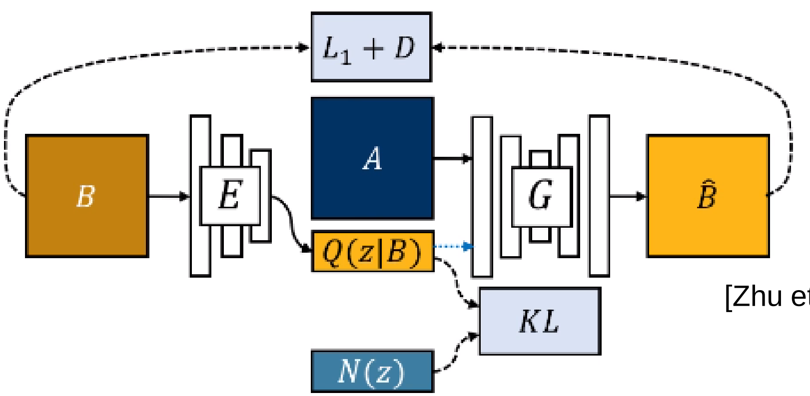
- One possible approach: pix2pix + noise



[Zhu et al. 2017]

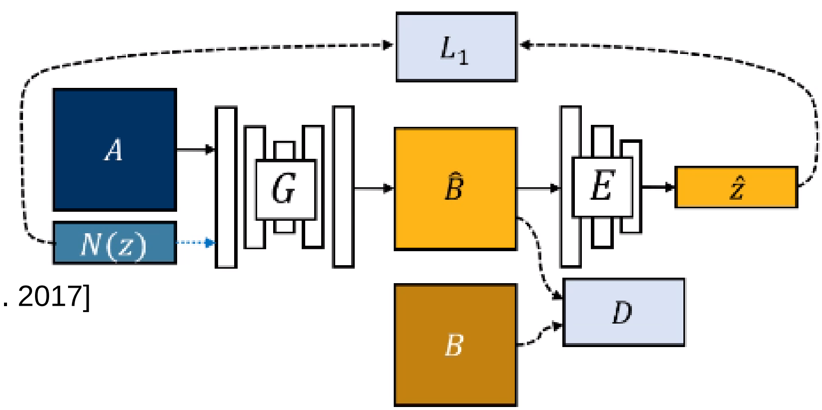
BicycleGAN: Multimodal I2I Translation

- Conditional Variational AutoEncoder GAN (cVAE-GAN)
  - The purpose is to reconstruct ground truth image
  - $B \rightarrow Z \rightarrow \hat{B}$
- Conditional Latent Regressor GAN (cLR-GAN)
  - The purpose is to recover the original latent vector
  - $Z \rightarrow B \rightarrow \hat{Z}$
- BicycleGAN: cVAE-GAN + cLR-GAN



Training cVAE-GAN

[Zhu et al. 2017]

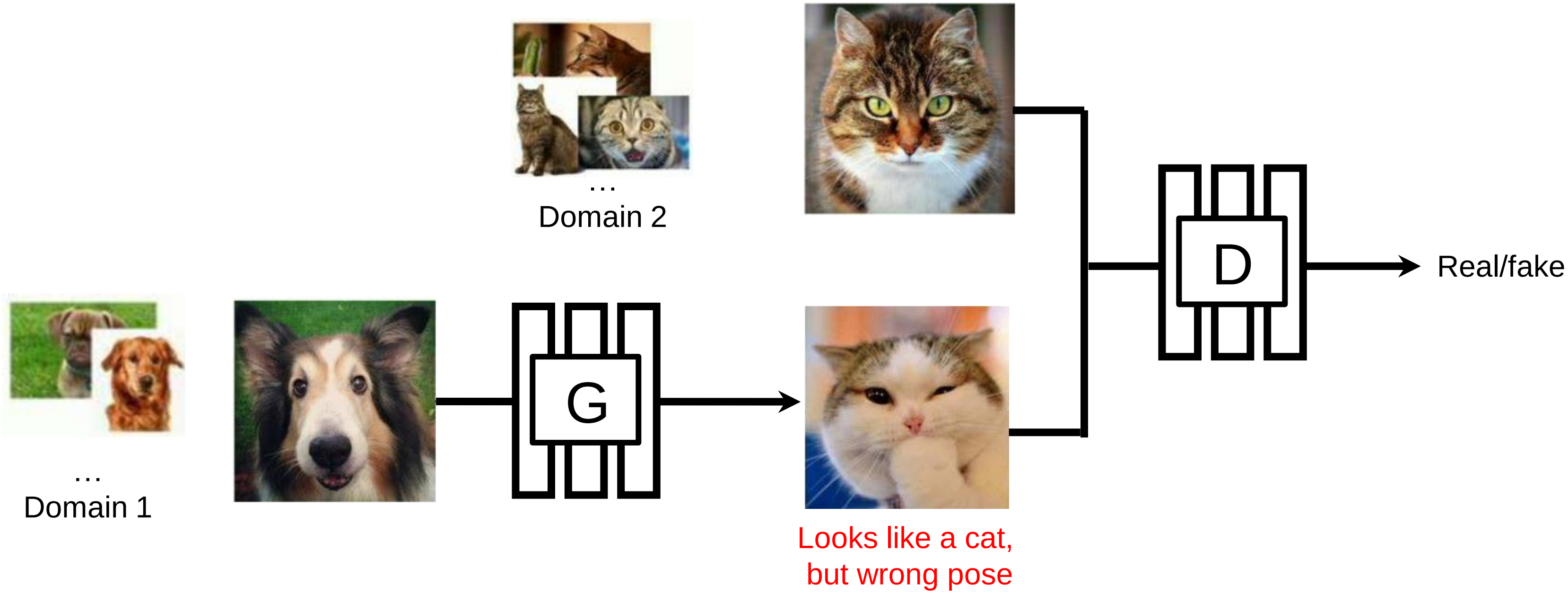


Training cLR-GAN

Unsupervised Image-to-image Translation

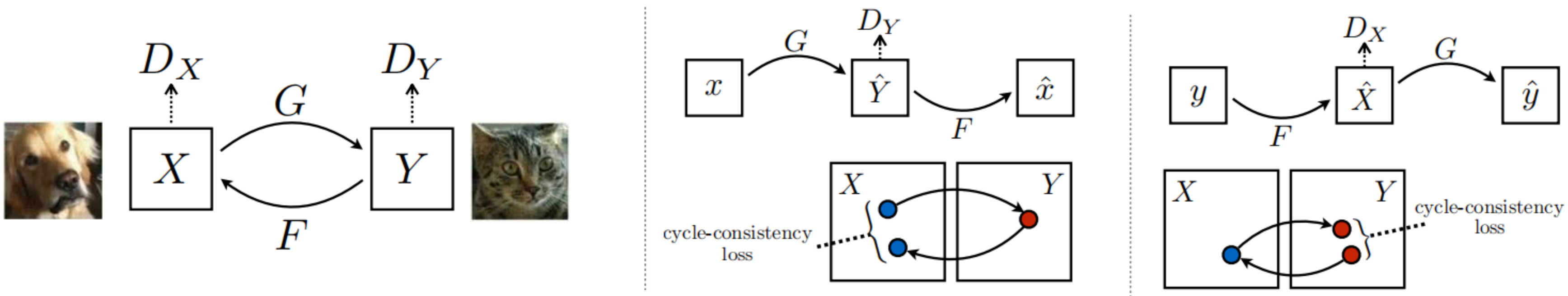
Plain GAN for unsupervised I2I translation

- Exploit supervision at **the level of sets**
  - Train a mapping $G: X \rightarrow Y$
  - Make the output $\hat{y} = G(x), x \in X$ indistinguishable from images $y \in Y$
  - Problem: individual inputs and outputs x and y are not paired up in a meaningful way.

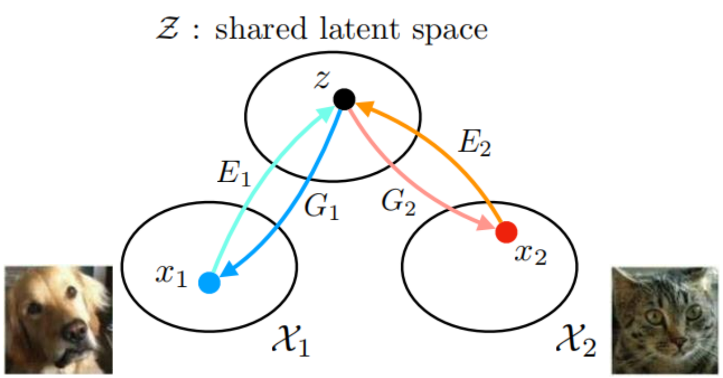


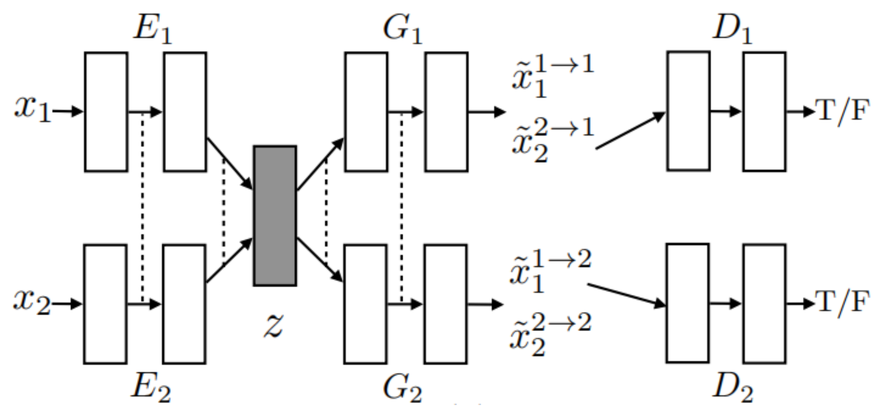
CycleGAN and UNIT

- CycleGAN: **Cycle-Consistent** Adversarial Networks [Zhu et al. 2017]
  - Cycle consistency



- UNIT: **UN**supervised **I**mage-to-image **T**ranslation [Liu et al. 2017]
  - Shared latent space





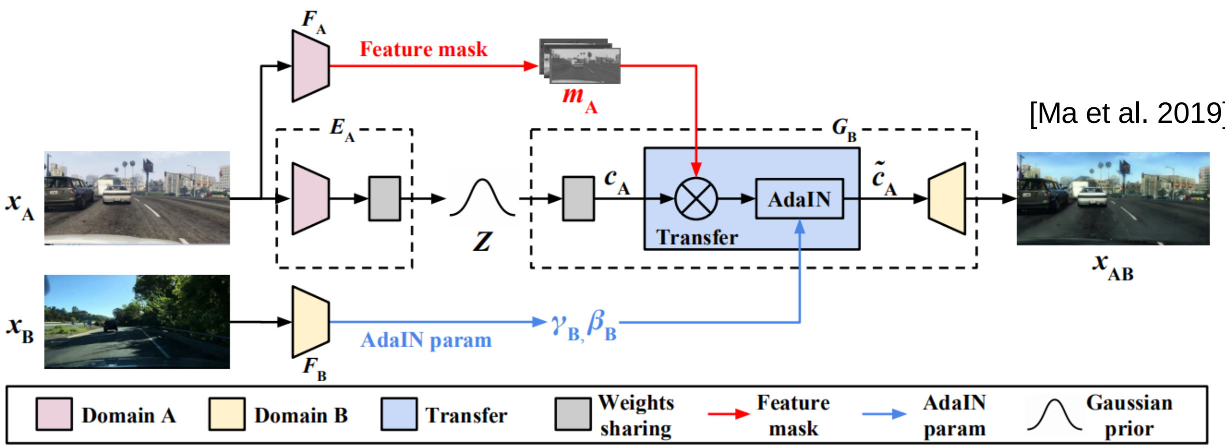
Multimodal unsupervised I2I Translation

- EGSC-IT: Exemplar Guided & Semantically Consistent Image-to-image Translation

  image → domain-shared **content** (e.g. objects' category, shape and spatial layout)

  image → domain-specific **style** (e.g. color and texture)

  - Weight sharing for domain-shared content -> *adopt* ***UNIT***
  - Exemplar-based AdaIN for domain-specific style
  - Feature masks for semantic consistency

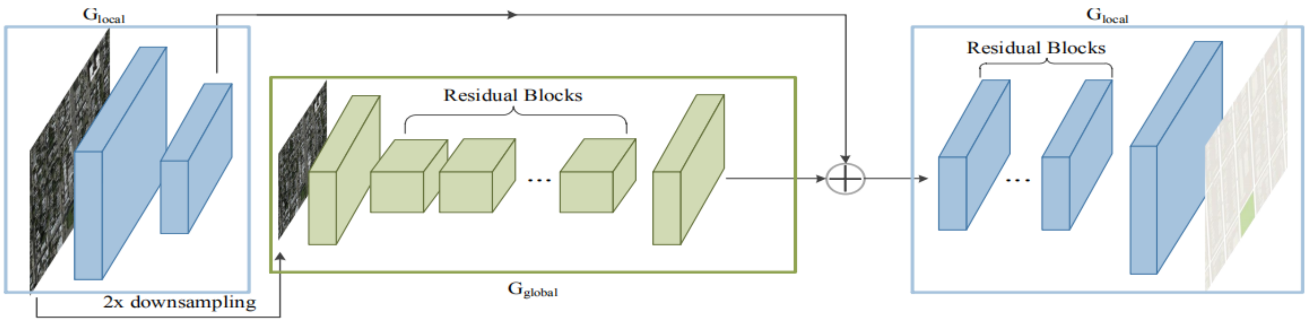


[Ma et al. 2019]

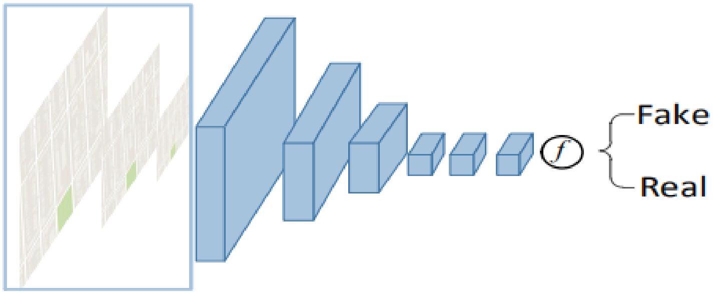
Image-to-image Translation in Geomatics

Aerial photos ⇄ maps

- CycleGAN-based model [Gu et al. 2019]
  - Generator (G):
    - G_{global} : a global generator operator at a low resolution
    - G_{local} : a local generator operator at a high resolution

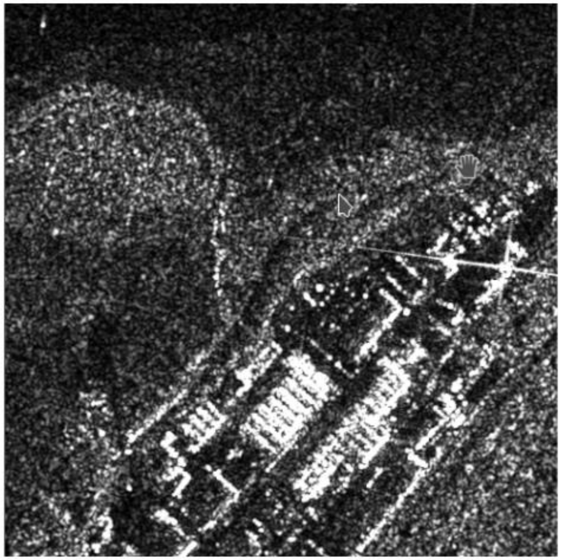


  - Discriminator (D): multi-scale discriminators



SAR image to optical image

- **Conditional GANs, CycleGAN, etc.**
- **Opportunity**
  - Reduce the speckle effect
  - Produce smooth textures
  - Fill missing content
  - …
- **Challenge**
  - Resolution
  - Level of detail
  - Fictional objects
  - …
- **Necessary optimization**
  - Size of the patches
  - Add several residual layers
  - Intensity clipping, normalization…

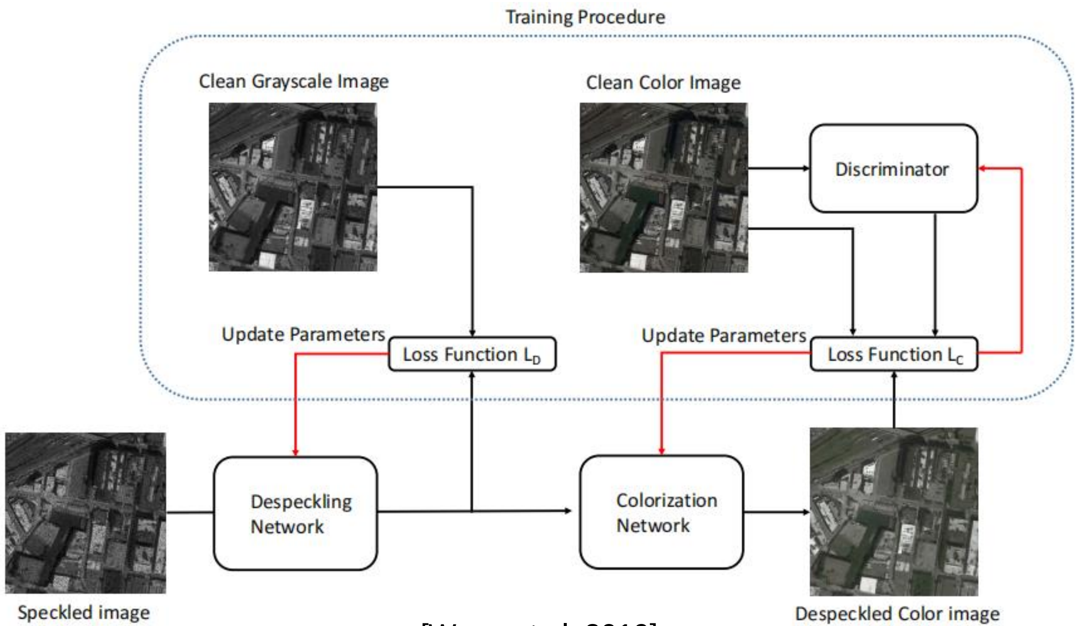


SAR image

Optical image

SAR image to optical image

- **SAR-GAN**
  - Learn a mapping from input speckled SAR images to visible images.
  - Despeckling sub-network G_D
  - Colorization subnetwork G_c
  - Generative adversarial loss



[Wang et al. 2019]

Summary

- **GAN**: a power tool for I2I translation

- **Representative works**
  - Supervised vs. unsupervised
  - Unimodal vs. multimodal
  - Most practical: unsupervised + multimodal

	Unsupervised	Multimodal
Pix2pix	×	×
BicycleGAN	×	√
CycleGAN	√	×
UNIT	√	×
EGSC-IT	√	√

- **I2I translation in geomatics**
  - Aerial photos to maps
  - SAR image to optical image

- **Challenges**
  - Mode collapse
  - Non-trivial training
  - Geometric changes

horse → zebra [Zhu et al. 2017]

Thank you