AIMS

African Institute for Mathematical Sciences - AMMI

29.08.2025

ARTICLE PRESENTATION AND IMPLEMENTATION

Image-to-Image Translation with Conditional Adversarial Networks

Phillip Isola Jun-Yan Zhu Tinghui Zhou Alexei A. Efros

Berkeley AI Research (BAIR) Laboratory, UC Berkeley

{isola,junyanz,tinghuiz,efros}@eecs.berkeley.edu

Presented By : **Audrey DJIOSSEU TIODO**

Academic Year 2024 - 2025

PRESENTATION OUTLINE

INTRODUCTION

Goal of the Paper

METHODOLOGY

EXPERIMENTS & RESULTS

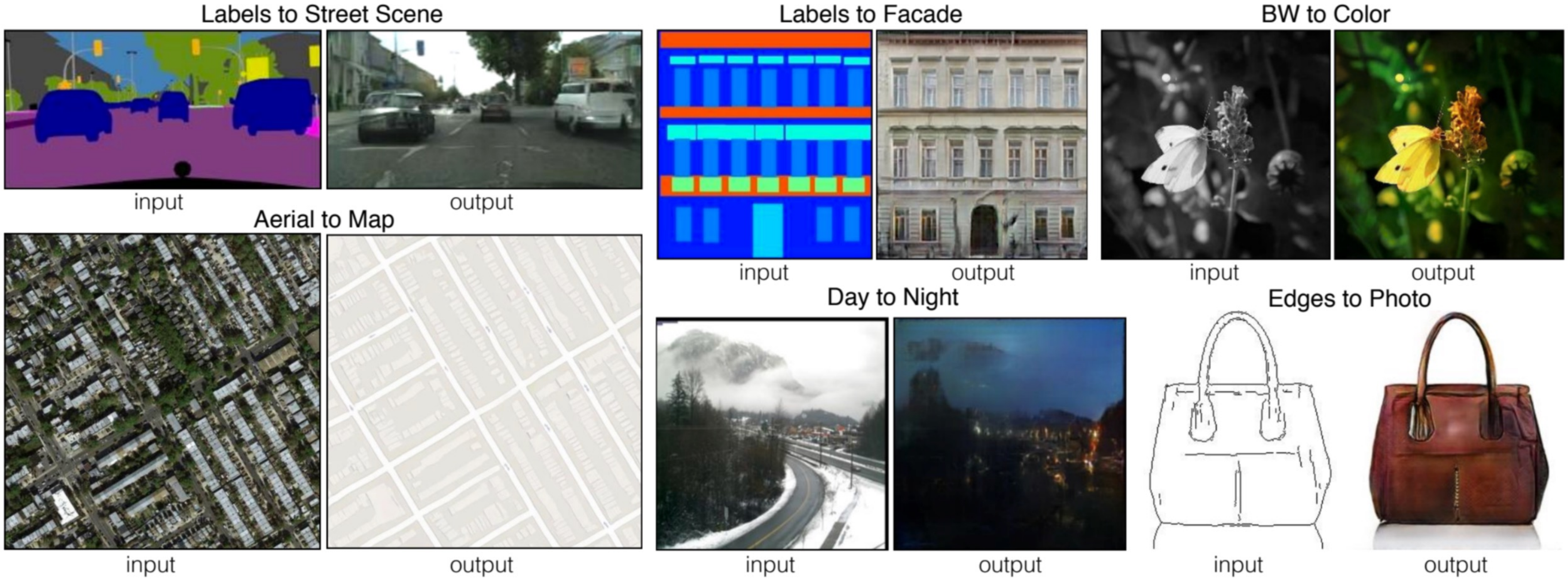
ANALYSIS

KEY CONTRIBUTIONS

CONCLUSION & REFERENCES

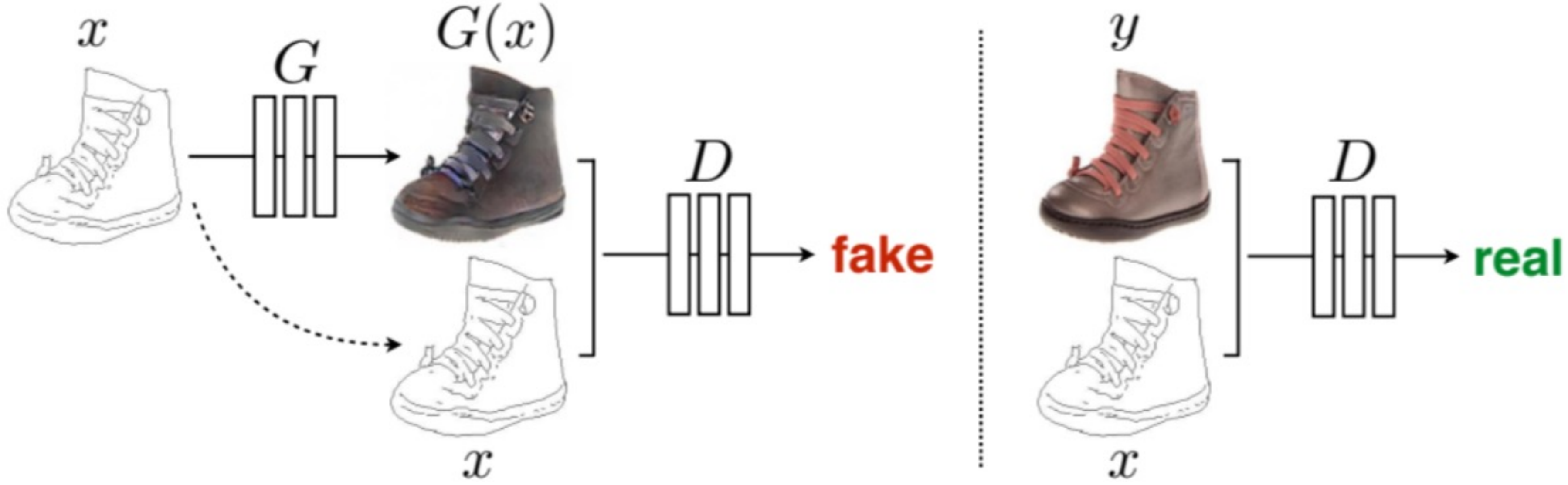
INTRODUCTION

1. Problem: translating an image from one domain to another (e.g., edges → photo, map → satellite, grayscale → color).



2. Limitation of traditional methods: each task required a hand-crafted algorithm and a specific loss function.

- Provide a **general framework** using **Conditional GANs (cGANs)** for diverse image-to-image translation tasks.



- Key idea: instead of manually designing losses, let the model learn the loss function directly.

METHODOLOGY

Conditional GAN (cGAN):

- ➢ Generator G(x,z): produces an image from the input.
  - ➢ Discriminator D(x,y): decides if the pair (input, output) is real or generated.

Objective function: adversarial loss + L1 distance → sharp and realistic outputs, reducing blurriness.

$$L_{cGAN}(G, D) = L_{GAN}(G, D) + \lambda L_1(G)$$

- $L_{GAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]$
- $L_1(G) = \mathbb{E}_{x,y,z}[||y - G(x, z)||_1]$

Architecture:

- ➢ Generator: **U-Net** with skip connections → preserves low-level details.
- ➢ Discriminator: **PatchGAN** → focuses on local image patches, enforcing texture realism.

EXPERIMENTS & RESULTS

Tested tasks:

- semantic labels → photo (Cityscapes, Facades),
- sketch/edges → photo,
- map ↔ aerial photos,
- grayscale → color,
- day → night,

Findings:

- cGAN produces **sharper and more realistic images** compared to L1/L2 losses.
- Perceptual tests (Amazon Mechanical Turk): generated images fooled humans about 20–25% of the time.
- Works well even with small datasets; efficient inference (< 1 sec per image on GPU).

ANALYSIS

Loss	Per-pixel acc.	Per-class acc.	Class IOU
Encoder-decoder (L1)	0.35	0.12	0.08
Encoder-decoder (L1+cGAN)	0.29	0.09	0.05
U-net (L1)	0.48	0.18	0.13
U-net (L1+cGAN)	**0.55**	**0.20**	**0.14**

- **Loss comparison**

L1 only → blurry.
L1 + cGAN → best balance of realism and accuracy.

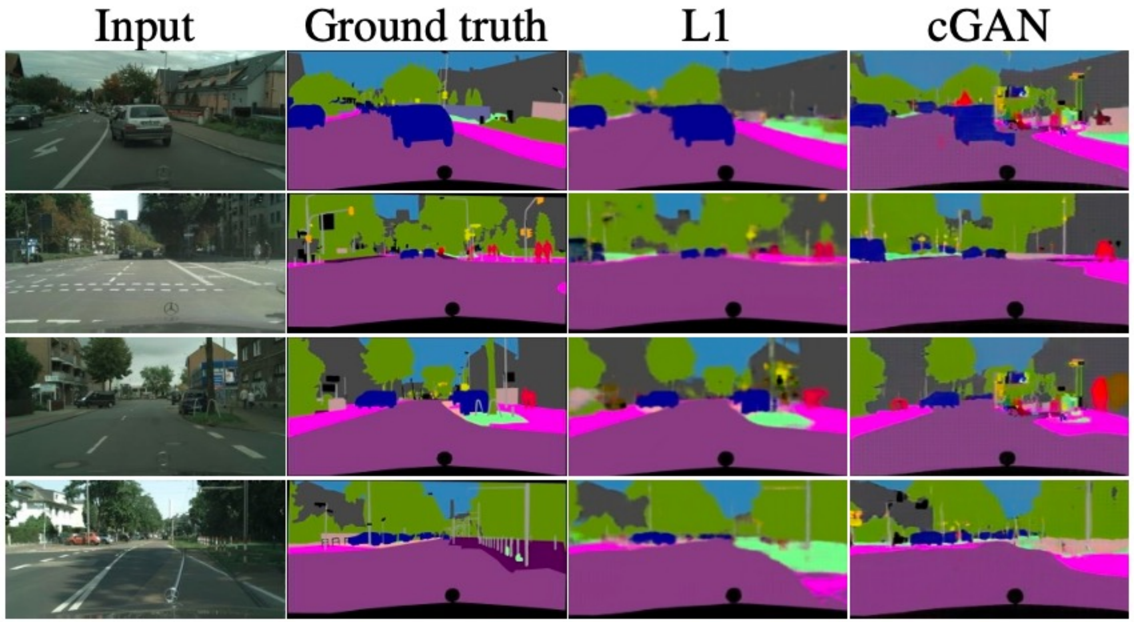
- **Patch size:** 70×70 PatchGAN gave the best trade-off between local realism and global consistency.
- **U-Net vs encoder-decoder:** U-Net significantly outperforms plain encoder-decoder networks.

KEY CONTRIBUTIONS

- A simple but **unified framework** for many image-to-image tasks.
- Showed that **learning the loss function** is more effective than hand-crafting it.
- Released **pix2pix code**, which became widely adopted in research and creative applications.

CONCLUSION

- Conditional GANs are a powerful solution for structured image-to-image translation.
- Limitations: for tasks like segmentation, simple L1 may still work better.
- Impact: inspired follow-up works such as **pix2pixHD, CycleGAN, and many more**.



REFERENCE

Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

AIMS