

# Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Indah Nurawaliah  
Muhammad Ikhsan Ilma  
Muhammad Farid Aqthar  
Sistem Informasi-Universitas Hasanuddin

## CONTENTS

|            |  |   |
|------------|--|---|
| <b>I</b>   | <b>Pendahuluan</b>                               | 1 |
| <b>II</b>  | <b>Rumusan Masalah</b>                           | 1 |
| <b>III</b> | <b>Tinjauan Pustaka</b>                          | 1 |
| III-A      | Generative Adversarial Networks (GANs) . . . . . | 1 |
| III-B      | Image-to-Image Translation . . . . .             | 1 |
| III-C      | Unpaired Image-to-Image Translation . . . . .    | 1 |
| III-D      | Cycle Consistency . . . . .                      | 2 |
| III-E      | Neural Style Transfer . . . . .                  | 2 |
| <b>IV</b>  | <b>Metode Penelitian</b>                         | 2 |
| IV-A       | Sumber Data . . . . .                            | 2 |
| IV-B       | Alur Kerja . . . . .                             | 2 |
| IV-B1      | Adversarial Loss . . . . .                       | 2 |
| IV-B2      | Cycle Consistency Loss . . . . .                 | 2 |
| IV-B3      | Full Objective . . . . .                         | 2 |
| <b>V</b>   | <b>Hasil dan Pembahasan</b>                      | 2 |
| V-A        | Style Transfer . . . . .                         | 2 |
| V-B        | Object Transfiguration . . . . .                 | 3 |
| V-C        | Season Transfer . . . . .                        | 3 |
| V-D        | Photograph Generation From Paintings . . . . .   | 3 |
| V-E        | Photograph Enhancement . . . . .                 | 3 |
| <b>VI</b>  | <b>Kesimpulan dan Saran</b>                      | 3 |
|            | <b>References</b>                                | 4 |

## LIST OF FIGURES

|   |           |   |
|---|-----------|---|
| 1 | . . . . . | 2 |
| 2 | . . . . . | 2 |
| 3 | . . . . . | 2 |
| 4 | . . . . . | 3 |
| 5 | . . . . . | 3 |
| 6 | . . . . . | 3 |
| 7 | . . . . . | 3 |
| 8 | . . . . . | 3 |

## LIST OF TABLES

# Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

**Abstract**—Salah satu teknik yang melibatkan model pelatihan otomatis dalam menerjemahkan gambar-ke-gambar adalah CycleGAN yang dapat modifikasi gambar yang terkontrol dan membutuhkan kumpulan data besar dari gambar berpasangan yang rumit untuk disiapkan atau terkadang tidak ada.

## I. PENDAHULUAN

Dalam terjemahan gambar-ke-gambar melibatkan pembuatan versi sintetis baru dari gambar tertentu dengan modifikasi tertentu, seperti menerjemahkan landscape musim panas ke musim dingin. Melatih model untuk terjemahan gambar-ke-gambar biasanya membutuhkan kumpulan data besar dari contoh berpasangan. Kumpulan data ini bisa sulit dan mahal untuk disiapkan, dan dalam beberapa kasus tidak mungkin, seperti foto lukisan oleh seniman yang sudah lama meninggal.

CycleGAN adalah teknik yang melibatkan model pelatihan otomatis dalam menerjemahkan gambar-ke-gambar tanpa contoh berpasangan. Model tersebut dilatih tanpa pengawasan menggunakan kumpulan gambar dari domain sumber dan target yang tidak perlu dikaitkan dengan cara apa pun.

## II. RUMUSAN MASALAH

Terjemahan Gambar-ke-Gambar melibatkan modifikasi gambar yang terkontrol dan membutuhkan kumpulan data besar dari gambar berpasangan yang rumit untuk disiapkan atau terkadang tidak ada. CycleGAN adalah teknik untuk melatih model terjemahan gambar tanpa pengawasan melalui arsitektur GAN menggunakan kumpulan gambar yang tidak berpasangan dari dua domain berbeda. CycleGAN telah didemonstrasikan pada berbagai aplikasi termasuk terjemahan musim, transformasi objek, transfer gaya, dan menghasilkan foto dari lukisan. Penerjemahan gambar-ke-gambar adalah tugas sintesis gambar yang membutuhkan pembuatan gambar baru yang merupakan modifikasi terkontrol dari gambar yang diberikan.

Tujuan dari Technical report ini adalah untuk memahami arsitektur GAN (Generative Adversarial Network) dan mempelajari cara mengimplementasikannya dengan menggunakan pytorch.

## III. TINJAUAN PUSTAKA

Convolution Neural Network (CNN) adalah perkembangan metode dari Multi Layer Perception (MLP). Namun, CNN memiliki jumlah dimensi yang lebih banyak dibandingkan dengan MLP. CNN memiliki masukan (input) array mulai dari dua dimensi hingga lebih. Pada laporan ini menggunakan beberapa arsitektur CNN yaitu inception/googleNet, resNet, dan denseNet.

### A. Generative Adversarial Networks (GANs)

telah mencapai hasil yang mengesankan dalam pembuatan gambar, pengeditan gambar, dan pembelajaran representasi. Metode terbaru mengadopsi ide yang sama untuk kondisional aplikasi pembuatan gambar, seperti text2image, image inpainting, dan prediksi masa depan, serta untuk domain lain seperti video dan data 3D. Kunci untuk Keberhasilan GAN adalah gagasan tentang kerugian permusuhan yang memaksa gambar yang dihasilkan pada prinsipnya tidak dapat dibedakan dari foto asli.

### B. Image-to-Image Translation

Gagasan terjemahan gambar-ke-gambar kembali setidaknya ke Hertzmann et al.'s Analogi Gambar, yang menggunakan model tekstur non-parametrik pada pasangan gambar pelatihan input-output tunggal. Pendekatan yang lebih baru menggunakan kumpulan data contoh input-output untuk mempelajari fungsi terjemahan parametrik menggunakan CNN. Pendekatan kami dibangun di atas kerangka "pix2pix" dari Isola et al. yang menggunakan generatif bersyarat jaringan musuh untuk mempelajari pemetaan dari input ke gambar keluaran. Ide serupa telah diterapkan pada berbagai tugas-tugas seperti menghasilkan foto dari sketsa atau dari atribut dan tata letak semantik. Namun, tidak seperti pekerjaan sebelumnya di atas, kami mempelajari pemetaan tanpa contoh berpasangan

### C. Unpaired Image-to-Image Translation

Beberapa metode yang lain juga menangani pengaturan yang tidak berpasangan, di mana tujuannya adalah untuk menghubungkan dua domain data:  $X$  dan  $Y$ . Rosales dkk. mengusulkan kerangka kerja Bayesian yang mencakup basis sebelumnya pada bidang acak Markov berbasis patch yang dihitung dari gambar sumber dan istilah kemungkinan yang diperoleh dari beberapa gambar gaya. Baru-baru ini, CoGAN dan lintas-modal jaringan adegan menggunakan strategi pembagian berat untuk belajar representasi umum di seluruh domain. Bersamaan dengan kami metode, Liu et al. memperluas kerangka kerja di atas dengan kombinasi autoencoder variasi dan jaringan permusuhan generatif. Baris lain bersamaan pekerjaan mendorong input dan output untuk berbagi fitur "konten" tertentu meskipun mungkin berbeda dalam "gaya". Metode ini juga menggunakan jaringan permusuhan, dengan istilah tambahan untuk menegakkan output agar dekat dengan input dalam ruang metrik yang telah ditentukan, seperti ruang label kelas, ruang piksel gambar, dan ruang fitur gambar.

#### D. Cycle Consistency

Gagasan menggunakan transitivitas sebagai cara untuk mengatur data terstruktur memiliki sejarah panjang. Di pelacakan visual, menegakkan konsistensi maju-mundur sederhana telah menjadi trik standar selama beberapa dekade. Dalam domain bahasa, memverifikasi dan meningkatkan terjemahan melalui “penerjemahan kembali dan rekonsiliasi” adalah sebuah teknik digunakan oleh penerjemah manusia, serta oleh mesin. Lagi baru-baru ini, konsistensi siklus tingkat tinggi telah digunakan dalam struktur dari gerakan, pencocokan bentuk 3D, cosegmentasi, keselarasan semantik padat, dan estimasi kedalaman. Dari jumlah tersebut, Zhou et al. dan Godard dkk. paling mirip dengan pekerjaan kami, karena mereka menggunakan kehilangan konsistensi siklus sebagai cara menggunakan transitivitas untuk mengawasi pelatihan CNN. Dalam karya ini, kami memperkenalkan kerugian serupa untuk mendorong G dan F agar konsisten dengan masing-masing lainnya. Bersamaan dengan pekerjaan kami, dalam proses yang sama, Yi et al. secara mandiri menggunakan tujuan yang sama untuk terjemahan gambar-ke-gambar yang tidak berpasangan, terinspirasi oleh dual belajar dalam terjemahan mesin

#### E. Neural Style Transfer

adalah cara lain untuk melakukan terjemahan gambar-ke-gambar, yang mensintesis gambar baru dengan menggabungkan konten satu gambar dengan gaya gambar lain (biasanya lukisan) berdasarkan mencocokkan statistik matriks Gram dari fitur-fitur dalam yang telah dilatih sebelumnya. Fokus utama kami, di sisi lain, adalah belajar pemetaan antara dua koleksi gambar, daripada menjadi tween dua gambar tertentu, dengan mencoba menangkap korespondensi antara struktur tampilan tingkat yang lebih tinggi. Oleh karena itu, metode kami dapat diterapkan untuk tugas-tugas lain, seperti lukisan → foto, transfigurasi objek, dll. di mana tunggal metode transfer sampel tidak bekerja dengan baik. Kami membandingkan dua metode ini di Bagian

### IV. METODE PENELITIAN

#### A. Sumber Data

Dalam technical report ini menggunakan dataset dari repository Github. Dataset yang kami gunakan pada penelitian ini adalah dataset gambar yang terdiri dari Cityscapes labelPhoto, Mapsaerial photograph, Architectural facades label-photo, Edges → shoes, HorseZebra and AppleOrange, SummerWinter Yosemite, PhotoArt for style transfer, Monet’s paintings → photos, Flower photo enhancement.

#### B. Alur Kerja

1) *Adversarial Loss*: Kami menerapkan Adversarial Loss untuk kedua fungsi pemetaan. Untuk fungsi pemetaan  $G : X \rightarrow Y$  dan diskriminatornya  $D_Y$  di mana  $G$  mencoba menghasilkan gambar  $G(x)$  yang terlihat mirip dengan citra dari domain  $Y$ , sedangkan  $D_Y$  bertujuan untuk membedakan antara sampel terjemahan  $G(x)$  dan sampel nyata  $y$ .  $G$  tujuan untuk meminimalkan tujuan ini melawan musuh  $D$  yang mencoba untuk

memaksimalkannya, yaitu  $\min_G \max_{D_Y} \mathcal{L}_{GAN}(G, D_Y, X, Y)$ . Kami memperkenalkan kerugian permusuhan serupa untuk pemetaan fungsi  $F : Y \rightarrow X$  dan diskriminatornya  $D_X$  juga: yaitu,  $\min_F \max_{D_X} \mathcal{L}_{GAN}(F, D_X, Y, X)$ . kami menyatakan tujuannya sebagai:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} [\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x)))]$$

Fig. 1.

2) *Cycle Consistency Loss*: Training secara teori, mempelajari pemetaan  $G$  dan  $F$  yang menghasilkan output yang didistribusikan secara identik sebagai target domain  $Y$  dan  $X$  masing-masing (tepatnya, ini membutuhkan  $G$  dan  $F$  menjadi fungsi stokastik). Namun, dengan kapasitas yang cukup besar, suatu jaringan dapat memetakan hal yang sama set gambar input ke permutasi acak gambar apa pun di domain target, di mana pemetaan yang dipelajari dapat menginduksi distribusi output yang sesuai dengan distribusi target. Dengan demikian, kerugian permusuhan saja tidak dapat menjamin bahwa fungsi yang dipelajari dapat memetakan input individu  $x_i$  ke keluaran yang diinginkan  $y_i$ .

3) *Full Objective*: Digunakan rumus

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F),$$

Fig. 2.

Dimana mengontrol kepentingan relatif dari dua tujuan. Kami bertujuan untuk memecahkan:

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y).$$

Fig. 3.

Perhatikan bahwa model kita dapat dilihat sebagai melatih dua “autoencoder” : kita mempelajari satu autoencoder  $F \circ G : X \rightarrow X$  bersama-sama dengan  $G \circ F$  lain :  $Y \rightarrow Y$ . Namun, au toencoder ini masing-masing memiliki struktur internal khusus: mereka memetakan gambar ke dirinya sendiri melalui representasi perantara yang adalah terjemahan dari gambar ke domain lain. Misalnya setup juga dapat dilihat sebagai kasus khusus “autoencoder permusuhan”, yang menggunakan kerugian permusuhan untuk melatih lapisan bottleneck autoencoder untuk mencocokkan distribusi target yang berubah-ubah. Dalam kasus kami, distribusi target untuk  $X \rightarrow X$  autoencoder adalah domain  $Y$

### V. HASIL DAN PEMBAHASAN

#### A. Style Transfer

Transfer gaya mengacu pada pembelajaran gaya artistik dari satu domain, seringkali lukisan, dan menerapkan gaya artistik

ke domain lain, seperti foto. CycleGAN didemonstrasikan dengan menerapkan gaya artistik dari Monet, Van Gogh, Cezanne, dan Ukiyo-e pada foto-foto landscape.

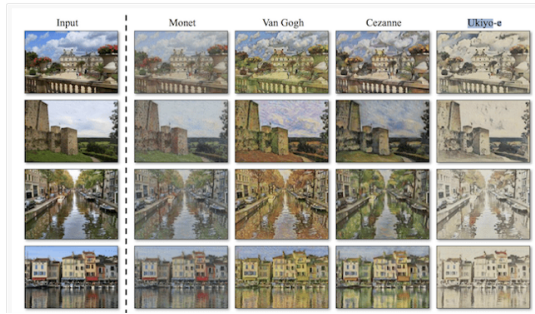


Fig. 4.

### B. Object Transfiguration

Transfigurasi objek mengacu pada transformasi objek dari satu kelas, seperti anjing ke kelas objek lain, seperti kucing. CycleGAN didemonstrasikan mengubah foto kuda menjadi zebra dan sebaliknya: foto zebra menjadi kuda. Jenis transfigurasi ini masuk akal karena kuda dan zebra terlihat serupa dalam ukuran dan struktur, kecuali warna mereka.

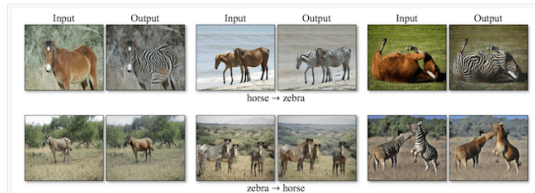


Fig. 5.

### C. Season Transfer

Transfer musim mengacu pada terjemahan foto yang diambil dalam satu musim, seperti musim panas, ke musim lain, seperti musim dingin. CycleGAN didemonstrasikan dalam menerjemahkan foto-foto lanskap musim dingin ke lanskap musim panas, dan kebalikan dari lanskap musim panas ke lanskap musim dingin.

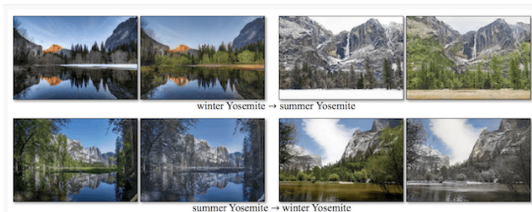


Fig. 6.

### D. Photograph Generation From Paintings

Generasi foto dari lukisan, seperti namanya, adalah sintesis gambar fotorealistik yang diberikan sebuah lukisan, biasanya oleh seniman terkenal atau adegan terkenal. CycleGAN didemonstrasikan dalam menerjemahkan banyak lukisan karya Monet menjadi foto-foto yang masuk akal.



Fig. 7.

### E. Photograph Enhancement

Peningkatan foto mengacu pada transformasi yang meningkatkan gambar asli dalam beberapa cara. CycleGAN ditunjukkan pada peningkatan foto dengan meningkatkan kedalaman bidang (misalnya memberikan efek makro) pada foto close-up bunga.



Fig. 8.

## VI. KESIMPULAN DAN SARAN

Terjemahan Gambar-ke-Gambar melibatkan modifikasi gambar yang terkontrol dan membutuhkan kumpulan data besar dari gambar berpasangan yang rumit untuk disiapkan atau terkadang tidak ada. CycleGAN adalah teknik untuk melatih model terjemahan gambar tanpa pengawasan melalui arsitektur

GAN menggunakan kumpulan gambar yang tidak berpasangan dari dua domain berbeda. CycleGAN telah didemonstrasikan pada berbagai aplikasi termasuk terjemahan musim, transfigurasi objek, transfer gaya, dan menghasilkan foto dari lukisan.

#### REFERENCES

- [1] Jun-Yan Zhu *et al* "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks" 2020 Berkeley AI Research (BAIR) laboratory, UC Berkeley 1703.10593v7: <http://arxiv.org/pdf/1703.10593.pdf>. [Accessed: Dec. 3, 2003].
- [2] L. Deng, O. Abdel-Hamid and D. Yu "A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, 2013, pp. 6669-6673. Available: <http://www.vmsk.org/Layman.pdf>. [Accessed: Dec. 3, 2003].
- [3] Hu, Wei, et al. "Deep convolutional neural networks for hyperspectral image classification." *Journal of Sensors* 2015 (2015).
- [4] W. Setiawan, M. . Utoyo, and R. Rulaningtyas, "Classification of neo-vascularization using convolutional neural network model," *TELKOMNIKA*, vol. 17, no. 1, pp. 463–473, 2019.
- [5] M. K. Hasan, Adiwijaya , S. A. Faraby "Klasifikasi Citra Multi-Kelas Menggunakan Convolutional Neural Network", *e-Proceeding of Engineering* : Vol.6, No.1 April 2019 — Page 2127, ISSN : 2355-9365.
- [6] S. I. Ahmad "arsitektur Convolutional Neural Network (CNN) Untuk Klasifikasi Jenis Dan Kesegaran Buah Pada Neraca Buah", Tesis – If185401, Available: [https://repository.its.ac.id/73567/1/05111850010020-Master\\_Thesis.pdf](https://repository.its.ac.id/73567/1/05111850010020-Master_Thesis.pdf)
- [7] Pardede Jasman, D. . Putra, " Implementasi DenseNet Untuk Mengidentifikasi Kanker Kulit Melanoma", *Jurnal Teknik Informatika dan Sistem Informasi* Volume 6 Nomor 3 Desember 2020, p-ISSN : 2443-2210, e-ISSN : 2443-2229