

Image-to-Image Translation with a Conditional Adversarial Networks (Pix2pix)

Andi Ario Ichsan Dharmawan (H071191031)

Computer Science
Hasanuddin University

CONTENTS

I	Introduction	1
II	Proposed Solutions	1
II-A	Generative Adversarial Network (GAN)	1
II-B	Architecture of GAN	1
II-B1	Generator	1
II-B2	Discriminator	2
II-C	Pix2pix GAN	2
III	Dataset	2
IV	Conclusion	2
	References	2

LIST OF FIGURES

1	Arsitektur GAN	1
2	Arsitektur Generator GAN	2
3	Arsitektur Discriminator	2
4	Arsitektur Pix2pix	2

LIST OF TABLES

Image-to-Image Translation with a Conditional Adversarial Networks (Pix2pix)

Abstract—Generative Adversarial Network adalah sebuah arsitektur jaringan saraf tiruan yang bertujuan untuk membentuk atau membangkitkan suatu data yang benar-benar baru, dari tidak ada menjadi ada. Pada perkembangan saat ini, GAN tidak hanya digunakan untuk mensintesis data dari suatu *noise*, tetapi telah mulai digunakan untuk mentranslasi data dari suatu domain ke data dengan domain berbeda. Beberapa penelitian yang telah dikembangkan antara lain, CycleGAN, Pix2pix, dan juga Domain Transfer Network. Implementasi model yang digunakan adalah Pix2pix pada dataset *facades*.

I. INTRODUCTION

Pada tahun 2014, Ian Goodfellow et al., menciptakan sebuah *framework* baru dalam model generative. *Framework* ini menggunakan pendekatan Adversarial Model yang dikenal *Generative Adversarial Network* (GAN). Generative Adversarial Networks (GAN) merupakan neural network yang digunakan untuk unsupervised learning. Ada banyak jenis implementasi GAN yang berbeda. Ada beberapa jenis pengaplikasian GAN antara lain pada video prediction, translasi tulisan menjadi citra, dan image to image translation.

Sampai saat ini, telah banyak *framework* yang mengadaptasi GAN dalam beberapa subyek penelitian, salah satunya adalah cross-domain atau domain translation atau domain transfer. Subyek ini fokus ke bagaimana GAN dapat mentransformasi suatu data dalam suatu domain ke dalam domain yang berbeda. salah satu kasus yang banyak dilakukan adalah *image-translation* atau translasi citra seperti transfer dari foto menjadi karikatur atau emoji. GAN mempelajari kerugian yang mencoba mengklasifikasikan apakah output berupa gambar nyata atau palsu. Karena GAN mempelajari kerugian yang beradaptasi dengan data, maka GAN dapat diterapkan di banyak tugas-tugas yang akan membutuhkan jenis yang sangat berbeda dari fungsi kerugian.

Dalam paper ini, akan ditunjukkan cara membangun dan melatih conditional generative adversarial network (cGAN). Sama seperti GAN mempelajari model data generatif, conditional GANs (cGAN) mempelajari model data conditional yang bertujuan sebagai translasi *image-to-image* dengan mengkondisikan gambar input dan menghasilkan gambar output yang sesuai.

Pix2pix merupakan model dari cGAN yang mempelajari pemetaan dari gambar masukan ke gambar keluaran. Pix2pix dapat diterapkan ke berbagai implementasi, termasuk mensintesis foto dari peta label, menghasilkan foto berwarna dari gambar hitam putih, mengubah foto Google Maps menjadi gambar udara, dan bahkan mengubah sketsa menjadi foto. Pada penelitian ini, akan dihasilkan gambar fasad bangunan dengan Database CMP Facade dengan menggunakan model pix2pix.

II. PROPOSED SOLUTIONS

A. Generative Adversarial Network (GAN)

Generative Adversarial Network atau GAN (Goodfellow, 2014) adalah model generative yang pendekatannya berdasarkan kepada differentiable generator network. GAN merupakan sebuah arsitektur jaringan saraf tiruan yang bertujuan untuk membentuk atau membangkitkan suatu data yang benar-benar baru, dari tidak ada menjadi ada. Secara singkat, jaringan GAN dilatih untuk mampu membangkitkan suatu gambar baru berdasarkan kumpulan gambar yang telah ia lihat sebelumnya selama proses pelatihan.

B. Architecture of GAN

Secara umum arsitektur GAN terdiri dari 2 jaringan yang disebut sebagai jaringan Generator dan Discriminator.

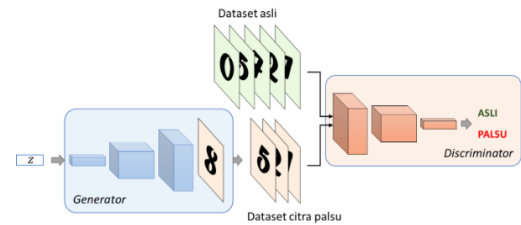


Fig. 1. Arsitektur GAN

1) *Generator*: Bentuk jaringan Generator dapat dilihat berkebalikan dengan struktur jaringan saraf pada umumnya. jaringan Generator menerima input sebuah vektor angka z , kemudian mengubahnya menjadi output gambar tiga dimensi. Vektor input z umumnya dibangkitkan secara acak, lalu dari angka sembarang tersebut Generator membangkitkan gambar yang juga sembarang.

Generator dari arsitektur UGATIT dapat dilihat pada Gambar 6. Terbagi oleh encoder, auxiliary classifier, dan decoder. Encoder downsampling memiliki dua layer konvolusi dengan strides berukuran dua, padding berukuran satu, sedangkan pada encoder bottleneck terdiri dari empat residual block. Bagian Auxiliary classifier dilatih untuk mempelajari bobot dari feature map pada domain asal menggunakan global average pooling dan global max pooling. bagian decoder bottleneck terdapat empat residual block dan bagian terakhir yaitu decoder up-sampling terdiri dari tiga layer konvolusi. Pada decoder up-sampling terakhir terdapat satu layer konvolusi dengan fungsi aktivasi Tanh. Penggunaan Tanh untuk menyamakan range dari tiap piksel citra $[-1, 1]$. Normalisasi yang diterapkan pada bagian encoder yaitu menggunakan Instance Normalization

(IN), dan AdaLIN pada decoder. Adapula fungsi aktivasi yang diterapkan pada jaringan generator ini yaitu ReLU dan tanh. Menurut junho kim et. al, fungsi normalisasi AdaLIN ini menggabungkan kelebihan dari Adaptive Instance Normalization dan Layer Normalization (LN). Adaptive Instance Normalization sendiri memiliki kemampuan untuk mengubah dan menyimpan konten gambar secara selektif.



Fig. 2. Arsitektur Generator GAN

2) *Discriminator*: Jaringan Discriminator merupakan jaringan klasifikasi biner yang menerima input gambar tiga dimensi dan mengeluarkan klasifikasi menyatakan input gambar adalah gambar asli dari dataset atau merupakan gambar buatan Generator. Discriminator dilatih dengan sekumpulan data yang dibangkitkan oleh Generator, dan sekumpulan data dari dataset, dan dilatih untuk bisa membedakan keduanya.

Berbeda dengan model generator, model ini terbagi oleh encoder, auxiliary classifier, dan classifier. Pada masing masing encoder menggunakan instance normalization, dan activation function Leaky ReLU. Model discriminator dan auxiliary classifier dilatih untuk membedakan apakah citra inputan adalah citra sintetis atau citra target. Pada discriminator menggunakan patch-GAN yang mengklasifikasi citra sintetis atau asli dengan ukuran 70x70 (lokal) dan 256x256 (global). Alur dari jaringan discriminator diawali dengan masukan citra sampel, dan keluaran dari jaringan discriminator apakah gambar masukan adalah gambar asli atau gambar palsu. Jaringan discriminator memanfaatkan attention feature maps yang menggunakan bobot dari encoded feature maps yang telah dilatih oleh auxiliary classifier. Menggunakan Spectral Normalization (SN) sebagai fungsi normalisasi. Pada discriminator global pada encoder downsampling berjumlah lebih banyak dari discriminator lokal, sebanyak enam layer sedangkan discriminator lokal hanya empat layer.

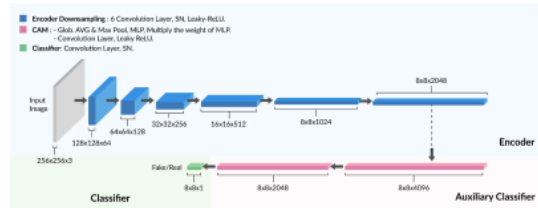


Fig. 3. Arsitektur Discriminator

C. Pix2pix GAN

Pix2Pix Generative Adversarial Network adalah pendekatan untuk melatih deep convolutional neural network untuk tugas

image-to-image translation. Pix2pix menunjukkan cara membangun dan melatih cGAN dengan mempelajari pemetaan dari gambar input ke gambar output. Pix2pix dapat diterapkan ke berbagai tugas, termasuk mensintesis foto dari peta label, menghasilkan foto berwarna dari gambar hitam putih, mengubah foto Google Maps menjadi gambar udara, dan bahkan mengubah sketsa menjadi foto.

Di pix2pix cGAN, gambar input akan dikondisikan dan menghasilkan gambar output yang sesuai. cGAN pertama kali diusulkan di Conditional Generative Adversarial Nets (Mirza dan Osindero, 2014)

Arsitektur jaringan akan berisi : 1. Generator dengan arsitektur berbasis U-Net 2. Diskriminator diwakili oleh pengklasifikasi PatchGAN convolutional.

(a) pix2pix

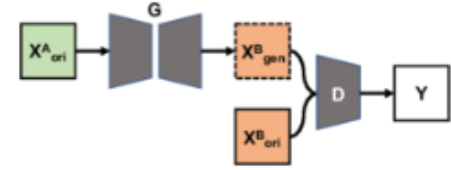


Fig. 4. Arsitektur Pix2pix

Pix2pix membutuhkan gambar pelatihan berpasangan yang disejajarkan dengan sempurna. Sebuah generator CNN dilatih untuk menghasilkan gambar yang mirip dengan gambar di domain B dari gambar di domain A dan diskriminator CNN dilatih secara bersamaan untuk membedakan gambar yang dihasilkan dari gambar nyata di domain B.

III. DATASET

Dataset yang digunakan adalah CMP Facade Database yang akan di generate dengan gambar bangunan yang diambil dari dataset tersebut. Dataset ini bertujuan untuk melakukan pengujian untuk mengetahui kinerja dari model.

IV. CONCLUSION

Hasil dalam makalah ini menunjukkan bahwa conditional adversarial network adalah pendekatan yang menjanjikan untuk banyak tugas terjemahan gambar-ke-gambar, terutama yang melibatkan output grafis yang sangat terstruktur. Jaringan ini mempelajari kerugian yang disesuaikan dengan tugas dan data yang ada, yang membuatnya dapat diterapkan di berbagai pengaturan.

REFERENCES

- [1] "pix2pix: Image-to-image translation with a conditional GAN" <https://learnopencv.com/paired-image-to-image-translation-pix2pix/>. Accessed on: June. 23, 2022.
- [2] S. Kida, <https://learnopencv.com/paired-image-to-image-translation-pix2pix/>. Accessed on: June. 23, 2022.. Accessed on: June. 23, 2022.

- [3] A. Sharma, "Pix2Pix:Image-to-Image Translation in PyTorch TensorFlow" <https://learnopencv.com/paired-image-to-image-translation-pix2pix/>. Accessed on: June. 23, 2022.
- [4] J. Lach, "SBFS: Steganography based file system," in *Proceedings of the 2008 1st International Conference on Information Technology, IT 2008, 19-21 May 2008, Gdansk, Poland*. Available: IEEE Xplore, <http://www.ieee.org>. [Accessed: 10 Sept. 2010].
- [5] "P 'Isola' Image-to-Image Translation with Conditional Adversarial Networks ," Nov. 26, 2018. Available: <https://arxiv.org/pdf/1611.07004.pdf>. [Accessed: June. 23, 2022].