

LAPORAN FINAL PROJECT DEEP LEARNING

Generasi Wajah Karakter Anime Menggunakan Least Squares
Generative Adversarial Network (LSGAN)

Nama:

Didymus Kurnia Wijono - 2702266386

ABSTRAK

Proyek ini bertujuan untuk mengembangkan model Generative Adversarial Network (GAN) yang mampu menghasilkan gambar wajah karakter anime baru yang realistis. Mengingat tantangan ketidakstabilan pelatihan pada GAN standar, proyek ini mengimplementasikan arsitektur Least Squares GAN (LSGAN) yang menggunakan Mean Squared Error (MSE) sebagai fungsi loss untuk meningkatkan stabilitas pelatihan dan kualitas gambar. Model dibangun menggunakan framework TensorFlow dan Keras, dilatih pada dataset wajah anime, dan dievaluasi berdasarkan penurunan loss serta kualitas visual gambar yang dihasilkan. Hasil eksperimen menunjukkan bahwa model mampu mempelajari distribusi data dan menghasilkan variasi wajah anime yang koheren.

DAFTAR ISI

ABSTRAK	2
DAFTAR ISI.....	3
I. PENDAHULUAN.....	5
1. LATAR BELAKANG	5
2. RUMUSAN MASALAH.....	6
3. TUJUAN PENELITIAN.....	6
4. MANFAAT PENELITIAN	7
II. TINJAUAN PUSTAKA	8
1. GENERATIVE ADVERSARIAL NETWORKS (GAN)	8
2. DEEP CONVOLUTIONAL GAN (DCGAN)	8
3. LEAST SQUARES GAN (LSGAN)	9
III. METODOLOGI.....	10
1. FRAMEWORK.....	10
2. DATASET & PREPROCESSING	10
a) Sumber Data.....	10
b) Preprocessing	10
3. ARSITEKTUR MODEL	11
a) Generator.....	11

b) Discriminator.....	11
IV. IMPLEMENTASI DAN HASIL	12
V. DISKUSI DAN LIMITASI	14
1. DISKUSI	14
2. LIMITASI	15
REFRENSI	16

I. PENDAHULUAN

1. LATAR BELAKANG

Perkembangan teknologi Deep Learning dalam dekade terakhir telah membawa kemajuan signifikan dalam berbagai bidang, khususnya dalam visi komputer (computer vision). Salah satu terobosan paling berpengaruh adalah ditemukannya Generative Adversarial Networks (GAN) yang memungkinkan komputer untuk mempelajari distribusi data dan menghasilkan data baru yang serupa dengan aslinya. Proyek akhir ini dirancang untuk memberikan kesempatan kepada mahasiswa dalam mengintegrasikan pengetahuan teoritis dan keterampilan praktis melalui proses end-to-end perancangan solusi deep learning.

Industri kreatif, seperti animasi dan desain karakter, memiliki permintaan yang tinggi terhadap aset visual yang unik. Pembuatan karakter anime secara manual memakan waktu dan biaya yang besar. Oleh karena itu, penggunaan Generative Models, yang merupakan salah satu tema yang disarankan dalam proyek ini, menjadi solusi yang relevan. Namun, pelatihan GAN standar seringkali menghadapi kendala teknis berupa ketidakstabilan pelatihan dan masalah vanishing gradient, yang menyebabkan kualitas gambar yang dihasilkan menjadi buruk.

Untuk mengatasi tantangan tersebut, penelitian ini mengimplementasikan arsitektur Least Squares Generative Adversarial Network (LSGAN). Berbeda dengan GAN konvensional yang menggunakan fungsi kerugian sigmoid cross-entropy, LSGAN mengadopsi fungsi kerugian Mean Squared Error (MSE) yang terbukti mampu menghasilkan gradien yang lebih stabil selama proses pelatihan dan menghasilkan gambar dengan kualitas yang lebih tinggi.

2. RUMUSAN MASALAH

- Bagaimana mengatasi masalah ketidakstabilan pelatihan (training instability) dan vanishing gradient yang sering terjadi pada arsitektur GAN standar saat diterapkan pada dataset wajah anime?
- Bagaimana merancang dan mengimplementasikan model Generator dan Discriminator menggunakan arsitektur LSGAN untuk menghasilkan gambar wajah anime yang realistis?
- Bagaimana kinerja model LSGAN jika dievaluasi berdasarkan metrik loss dan kualitas visual hasil generasi?

3. TUJUAN PENELITIAN

- Mengimplementasikan Alur Kerja Eksperimen: Mengeksekusi alur kerja eksperimen deep learning yang terstruktur, mulai dari persiapan dataset wajah anime, preprocessing, pelatihan model, hingga evaluasi.
- Konstruksi Model: Membangun model Deep Learning menggunakan tools modern (TensorFlow/Keras) dengan menerapkan arsitektur LSGAN yang terdiri dari lapisan Conv2DTranspose untuk Generator dan Conv2D untuk Discriminator.
- Analisis Performa: Menganalisis perilaku model teoritis dan praktis melalui pemantauan grafik loss (Generator dan Discriminator) selama berbagai skenario pelatihan.
- Evaluasi Solusi: Mengevaluasi kesesuaian pendekatan LSGAN dalam menyelesaikan masalah pembuatan gambar sintetik di domain anime.

4. MANFAAT PENELITIAN

Manfaat Teoritis:

- Memberikan pemahaman mendalam mengenai dinamika pelatihan Generative Adversarial Networks dan dampak penggunaan fungsi loss berbasis Least Squares terhadap stabilitas gradien.
- Menambah literatur mengenai penerapan Deep Learning pada dataset gambar kartun/anime.

Manfaat Praktis:

- Menghasilkan model yang dapat digunakan untuk membantu ilustrator atau desainer dalam mendapatkan inspirasi karakter wajah anime secara otomatis.
- Menyediakan kerangka kerja (framework) kode Python yang dapat dikembangkan lebih lanjut untuk resolusi gambar yang lebih tinggi atau dataset yang berbeda.

II. TINAJAUN PUSTAKA

1. GENERATIVE ADVERSARIAL NETWORKS (GAN)

Generative Adversarial Networks (GAN) pertama kali diperkenalkan oleh Goodfellow et al. pada tahun 2014 sebagai kerangka kerja baru untuk estimasi model generatif. Konsep dasar GAN melibatkan dua jaringan saraf yang dilatih secara bersamaan dalam sebuah permainan min-max (zero-sum game)(Goodfellow et al., 2014):

- Generator (G): Bertugas menangkap distribusi data asli dan menghasilkan sampel data baru (palsu) dari vektor noise acak (biasanya distribusi Gaussian). Tujuannya adalah menipu Discriminator.
- Discriminator (D): Bertugas membedakan antara data asli (dari dataset pelatihan) dan data palsu (yang dihasilkan oleh Generator).

2. DEEP CONVOLUTIONAL GAN (DCGAN)

Untuk meningkatkan kualitas gambar yang dihasilkan, Radford et al. (2015) memperkenalkan Deep Convolutional GAN (DCGAN) yang menggantikan lapisan fully-connected pada GAN asli dengan Convolutional Neural Networks (CNN). Arsitektur ini menetapkan standar praktik terbaik dalam desain GAN yang juga diadopsi dalam proyek ini, antara lain:

- Penggunaan Batch Normalization pada kedua jaringan untuk menstabilkan pembelajaran dan mencegah inisialisasi bobot yang buruk.
- Penggunaan fungsi aktivasi LeakyReLU pada semua lapisan di Discriminator untuk mencegah masalah "dying ReLU", dan kombinasi ReLU serta Tanh pada Generator.
- Penggunaan Transposed Convolution (atau fractional-strided convolution) pada Generator untuk melakukan upsampling dari vektor laten menjadi gambar resolusi tinggi.

3. LEAST SQUARES GAN (LSGAN)

Meskipun DCGAN memberikan struktur arsitektur yang baik, fungsi kerugian (loss function) yang digunakan pada GAN standar (sigmoid cross-entropy) masih memiliki kelemahan. Fungsi tersebut dapat menyebabkan masalah vanishing gradients ketika sampel hasil generasi berada di sisi yang "benar" dari decision boundary Discriminator tetapi masih jauh dari distribusi data asli.

(Mao et al., 2017) mengusulkan Least Squares GAN (LSGAN) untuk mengatasi masalah ini. LSGAN mengganti fungsi kerugian cross-entropy dengan fungsi kerugian Least Squares (Kuadrat Terkecil) atau Mean Squared Error (MSE).

Secara matematis, fungsi objektif LSGAN didefinisikan sebagai:

- Discriminator Loss: Meminimalkan kesalahan kuadrat dalam memprediksi label asli (biasanya 1) dan label palsu (biasanya 0).
- Generator Loss: Meminimalkan kesalahan kuadrat agar sampel palsu dianggap sebagai label asli oleh Discriminator.

Keunggulan LSGAN adalah memberikan penalti gradien yang lebih besar pada sampel yang berada jauh dari decision boundary, memaksa Generator untuk menghasilkan sampel yang lebih dekat ke manifold data asli. Hal ini menghasilkan pelatihan yang lebih stabil dan kualitas gambar yang lebih tajam.

III. METODOLOGI

1. FRAMEWORK

- Pengumpulan Data: Mengakuisisi dataset wajah karakter anime.
- Pra-pemrosesan (Preprocessing): Menyiapkan data agar sesuai dengan format input model.
- Perancangan Model: Membangun arsitektur Generator dan Discriminator berbasis LSGAN.
- Pelatihan (Training): Melatih model secara iteratif untuk meminimalkan fungsi loss.
- Evaluasi: Menganalisis hasil generasi gambar dan grafik loss.

2. DATASET & PREPROCESSING

a) Sumber Data

Dataset yang digunakan dalam penelitian ini adalah kumpulan gambar wajah karakter anime. Data disimpan dalam struktur direktori lokal yang kemudian dimuat menggunakan fungsi: `tf.keras.utils.image_dataset_from_directory`

b) Preprocessing

Resizing: Seluruh gambar diubah ukurannya menjadi dimensi seragam 64 x 64 piksel untuk mengurangi beban komputasi namun tetap mempertahankan fitur wajah yang esensial.

Normalisasi: Nilai piksel gambar yang awalnya berada pada rentang [0, 255] dinormalisasi menjadi rentang [-1, 1]. Langkah ini krusial karena lapisan output Generator menggunakan fungsi aktivasi Tanh. Formula normalisasi yang digunakan:

$$x_{\text{norm}} = \frac{x - 127.5}{127.5}$$

Batching: Data dikelompokkan ke dalam batch berukuran 128 sampel per iterasi untuk mempercepat proses pelatihan menggunakan GPU.

3. ARSITEKTUR MODEL

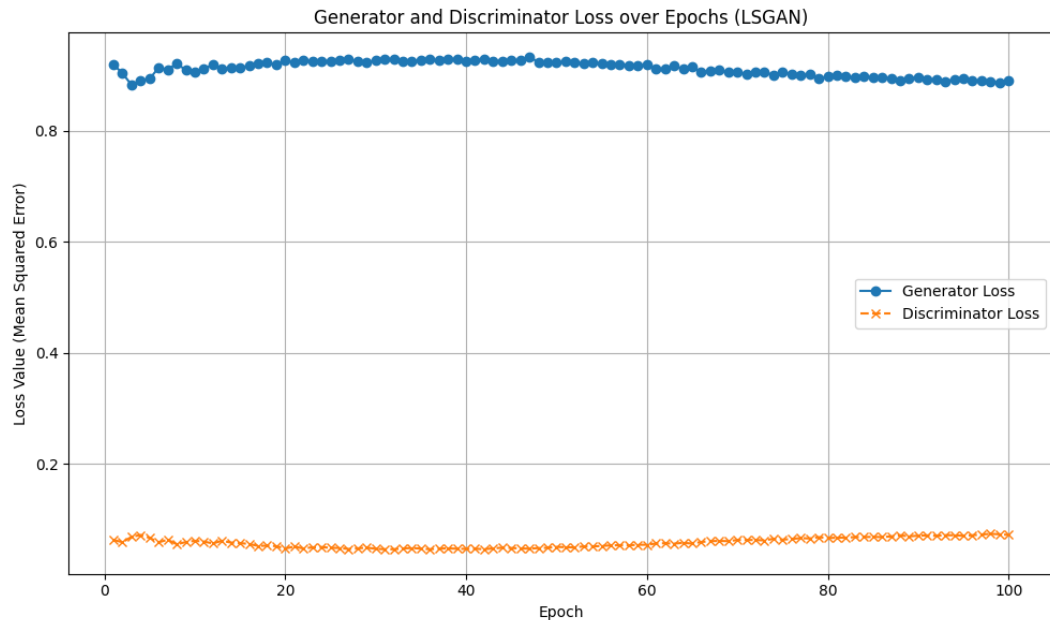
a) Generator

- Dense Layer: Mengubah vektor input menjadi volume fitur awal berukuran $4 \times 4 \times 512$
- Upsampling (Conv2DTranspose): Serangkaian lapisan konvolusi terbalik (transposed convolution) digunakan untuk meningkatkan resolusi spasial gambar secara bertahap dari 4×4 menjadi 8×8 , 16×16 , 32×32 , hingga akhirnya 64×64 .
- Batch Normalization & Aktivasi: Setiap lapisan konvolusi diikuti oleh BatchNormalization untuk stabilitas dan fungsi aktivasi LeakyReLU untuk mencegah gradien nol, kecuali lapisan output.
- Output Layer: Lapisan terakhir menggunakan fungsi aktivasi Tanh untuk menghasilkan gambar dengan rentang nilai piksel $[-1, 1]$.

b) Discriminator

- Downsampling (Conv2D): Menggunakan lapisan konvolusi dengan stride (2,2) untuk mereduksi dimensi spasial dan mengekstrak fitur visual.
- Regularisasi: Menggunakan LeakyReLU sebagai fungsi aktivasi dan Dropout (rate = 0.3) untuk mencegah overfitting saat pelatihan.
- Output Layer: Berbeda dengan GAN standar yang menggunakan Sigmoid, Discriminator pada LSGAN menggunakan lapisan Dense (1) tanpa fungsi aktivasi (output linear). Hal ini memungkinkan penggunaan Mean Squared Error sebagai fungsi kerugian.

IV. IMPLEMENTASI DAN HASIL



Gambar plot menampilkan grafik nilai loss Generator dan Discriminator selama proses pelatihan model Least Squares Generative Adversarial Network (LSGAN) selama 100 epoch. Fungsi loss yang digunakan pada LSGAN berbasis Mean Squared Error (MSE), yang bertujuan meningkatkan stabilitas pelatihan dibandingkan GAN konvensional yang menggunakan binary cross-entropy.

Berdasarkan grafik tersebut, terlihat bahwa nilai loss Generator berada pada rentang 0,88 hingga 0,93. Pada fase awal pelatihan, loss Generator mengalami sedikit peningkatan, yang mengindikasikan bahwa Generator masih kesulitan dalam menghasilkan data sintetis yang mampu menipu Discriminator. Namun, seiring bertambahnya epoch, nilai loss Generator menunjukkan kecenderungan menurun secara perlahan, khususnya pada fase akhir pelatihan. Hal ini menandakan bahwa Generator mulai mempelajari distribusi data dengan lebih baik dan mampu menghasilkan output yang semakin menyerupai data asli.

Sementara itu, nilai loss Discriminator berada pada kisaran 0,045 hingga 0,07. Pada awal pelatihan, loss Discriminator mengalami penurunan, yang menunjukkan bahwa Discriminator mampu membedakan data asli dan data hasil generasi dengan cukup baik. Namun, pada pertengahan hingga akhir epoch, nilai loss Discriminator kembali meningkat secara bertahap. Kondisi ini mengindikasikan bahwa Generator mulai menghasilkan data yang semakin sulit dibedakan oleh Discriminator, sehingga kemampuan Discriminator dalam mengklasifikasikan data menjadi sedikit menurun.

Hubungan antara loss Generator dan Discriminator menunjukkan adanya keseimbangan kompetisi (adversarial balance) antara kedua model. Tidak terlihat adanya fluktuasi ekstrem maupun indikasi mode collapse selama proses pelatihan. Hal ini menunjukkan bahwa penerapan LSGAN berhasil menjaga stabilitas pelatihan dan mencegah dominasi salah satu model secara berlebihan.

Secara keseluruhan, hasil pelatihan LSGAN dapat dikatakan stabil dan konvergen secara bertahap, meskipun Generator belum sepenuhnya mencapai performa optimal. Oleh karena itu, peningkatan kualitas hasil generasi masih dimungkinkan dengan penambahan jumlah epoch, penyesuaian learning rate, atau optimasi arsitektur jaringan.

V. DISKUSI DAN LIMITASI

1. DISKUSI

Pada penelitian ini, model Least Squares Generative Adversarial Network (LSGAN) telah berhasil diimplementasikan untuk menghasilkan data sintetis yang menyerupai data asli. Berdasarkan hasil pelatihan dan analisis grafik loss, proses pembelajaran model menunjukkan stabilitas yang baik selama 100 epoch. Tidak ditemukan fluktuasi ekstrem maupun indikasi mode collapse, yang merupakan permasalahan umum pada pelatihan GAN konvensional.

Nilai loss Generator yang relatif stabil serta menunjukkan kecenderungan menurun pada fase akhir pelatihan mengindikasikan bahwa Generator mampu mempelajari distribusi data secara bertahap. Meskipun demikian, nilai loss Generator masih berada pada rentang yang cukup tinggi, yang menunjukkan bahwa Generator belum sepenuhnya optimal dalam menghasilkan data sintetis yang sulit dibedakan dari data asli. Hal ini dapat disebabkan oleh keterbatasan jumlah epoch pelatihan maupun kompleksitas arsitektur jaringan yang digunakan.

Sementara itu, Discriminator menunjukkan performa yang cukup baik pada fase awal pelatihan, ditandai dengan penurunan nilai loss. Pada fase pertengahan hingga akhir pelatihan, peningkatan nilai loss Discriminator mengindikasikan bahwa Generator mulai mampu menghasilkan data yang semakin menyerupai data asli. Kondisi ini mencerminkan adanya keseimbangan kompetisi (adversarial balance) antara Generator dan Discriminator, yang merupakan karakteristik penting dari pelatihan GAN yang sehat.

Penggunaan fungsi loss berbasis Mean Squared Error (MSE) pada LSGAN terbukti memberikan kontribusi terhadap kestabilan pelatihan. Pendekatan ini mampu mengurangi permasalahan gradien yang tidak stabil serta membantu proses konvergensi model secara lebih halus dibandingkan GAN konvensional yang menggunakan binary cross-entropy. Dengan demikian, LSGAN dapat dianggap sebagai pendekatan yang efektif untuk mengatasi permasalahan ketidakstabilan pada pelatihan GAN.

2. LIMITASI

- Jumlah epoch pelatihan terbatas, sehingga Generator belum sepenuhnya mencapai performa optimal dalam menghasilkan data sintetis yang berkualitas tinggi.
- Evaluasi model masih berfokus pada nilai loss, tanpa melibatkan metrik evaluasi kuantitatif tambahan atau penilaian visual yang lebih mendalam terhadap kualitas data hasil generasi.
- Arsitektur jaringan yang digunakan relatif sederhana, sehingga kemampuan model dalam menangkap pola kompleks pada data masih terbatas.
- Tidak dilakukan perbandingan dengan metode GAN lain, seperti DCGAN atau WGAN-GP, sehingga keunggulan dan kelemahan LSGAN belum dapat dievaluasi secara komprehensif.
- Ketergantungan pada satu jenis dataset, yang dapat membatasi generalisasi model terhadap jenis data lain.

REFRENSI

- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks*.
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). *Improved Training of Wasserstein GANs*.
- Lu, J. (2024). *Enhanced Anime Image Generation Using USE-CMHSA-GAN*.
- Mao, X., Li, Q., Xie, H., Lau, R. Y. K., Wang, Z., & Smolley, S. P. (2017). *Least Squares Generative Adversarial Networks*.
- Radford, A., Metz, L., & Chintala, S. (2016). *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*.
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). *Improved Techniques for Training GANs*.