

**Aisyah Journal of Informatics and Electrical Engineering**  
Universitas Aisyah Pringsewu



Journal Homepage  
<http://jti.aisyahuniversity.ac.id/index.php/AJIEE>



---

**PEMBUATAN GAMBAR SINTESES DARI DEKSRIPSI TEKS DENGAN  
ALGORITMA GENERATIVE ADVERSARIAL NETWORK**

**Rakhmi Khalida<sup>1</sup>, R Wisnu Prio Pamungkas<sup>2</sup>, Siti Setiawati<sup>3</sup>**

<sup>1,2,3</sup>Teknik Informatika, Fakultas Teknik  
Universitas Bhayangkara Jakarta Raya

[rakhmi.khalida@dsn.ubharajaya.ac.id](mailto:rakhmi.khalida@dsn.ubharajaya.ac.id)<sup>1</sup>, [wisnu.prio@dsn.ubharajaya.ac.id](mailto:wisnu.prio@dsn.ubharajaya.ac.id)<sup>2</sup>,  
[siti.setiawati@dsn.ubharajaya.ac.id](mailto:siti.setiawati@dsn.ubharajaya.ac.id)<sup>3</sup>

**ABSTRACT**

Recently computers have been able to produce realistic photos from text. This is one of the potentials of machine learning to be used creatively. Machine learning is the field of solving problems that require an equivalent understanding of human intelligence. In this study using the Generative Adversarial Networks (GAN) algorithm is used to create images from text descriptions. The basic GAN architecture consists of 2 networks called a Generator and Discriminator network. The results of this study is images that are still not detailed in interpreting a text description, but the authors try to produce images that inspire, images can be more poetic when tried using poetry, lyrics, or book quotes.

**Keywords:** *GAN, Image Synthesis, Text Description*

**ABSTRAK**

Baru-baru ini komputer mampu menghasilkan foto-foto yang realistis dari sebuah teks. Hal ini adalah salah satu potensi dari machine learning untuk digunakan secara kreatif. Machine learning adalah bidang menyelesaikan masalah-masalah yang membutuhkan pemahaman yang setara dengan kecerdasan manusia. Pada penelitian ini menggunakan algoritme Generative Adversarial Networks (GAN) digunakan untuk menciptakan gambar dari deskripsi teks. Dasar arsitektur GAN terdiri dari 2 jaringan yang disebut sebagai jaringan Generator dan Discriminator. Hasil dari penelitian ini berupa gambar yang masih tidak detail dalam memaknai sebuah deskripsi teks, tetapi penulis mencoba menghasilkan gambar yang menginspirasi, gambar dapat lebih puitis ketika dicoba menggunakan puisi, lirik, atau kutipan buku.

**Kata Kunci:** *GAN, Sintesis Gambar, Deskripsi Teks*

## I. PENDAHULUAN

Baru-baru ini komputer mampu menghasilkan foto-foto yang realistis dari sebuah deskripsi teks. Hal ini dikenal dengan sebutan sintesis teks ke gambar. Sintesis teks ke gambar adalah menerjemahkan sebuah teks atau kalimat yang diinginkan ke dalam sebuah gambar yang digenerasi oleh komputer secara otomatis. Sintesis teks ke gambar adalah bagian dari model generatif yang telah menjadi aplikasi penting dalam *computer vision*. *Computer vision* adalah bagian dari *machine learning*.

Sejumlah peneliti berhasil melatih algoritme yang dapat menerjemahkan penjelasan seseorang dalam bentuk teks atau dalam bentuk kalimat menjadi foto atau gambar. Metode yang disebut generative adversarial network (GAN) digunakan untuk mengeksplorasi seni dengan *machine learning*. Algoritme Generative Adversarial Network (GAN), dirilis pada Januari oleh peneliti dari Deep Learning Technology Center Microsoft.

Penelitian yang dilakukan [1] (Tao Xu) dengan algoritme GAN tersebut mampu menciptakan gambar yang hampir mirip burung. Hal ini dapat terjadi ketika peneliti melatihnya pada dataset tertentu, misalnya gambar burung, tetapi peneliti masih gagal melakukannya saat memberikan dataset yang jumlahnya lebih besar dan beragam. Menurut peneliti dari Microsoft [1] cara kerja dari algoritme GAN adalah "dataset yang terdiri dari gambar dan keterangan, yang memungkinkan mesin ini untuk berlatih menyesuaikan kata-kata dengan gambar visualnya. Misalnya, GAN bisa menciptakan gambar burung ketika keterangannya menyebutkan kata burung, dan memahami seperti apa gambar burung sebenarnya".

Hal ini menarik, menantang dan bermanfaat. Tujuan utama penelitian ini adalah mencoba sebuah kecerdasan buatan yang mampu menciptakan gambar dari deskripsi teks sederhana menjadi sebuah foto, yang banyak dapat dimanfaatkan oleh aplikasi edit foto, bidang fashion desain juga arsitek bangunan bahkan pelukis. Selain untuk menghilangkan bosan, menciptakan gambar dari deskripsi teks juga bisa digunakan sebagai alat praktis.

Misalnya, proyek percobaan ini dapat dimanfaatkan untuk menciptakan tokoh fiktif acara TV, film, dan animasi. Penelitian ini adalah hasil orisinal karya penulis juga belum pernah diterbitkan dimanapun dan pengembangan dari penelitian ini masih terbuka lebar.

## II. TINJAUAN PUSTAKA

### A. Generative Adversarial Networks (GAN) [2]

Gagasan utama di balik Generative Adversarial Networks (GAN) [2] adalah mempelajari dua jaringan - jaringan Generator G yang mencoba menghasilkan gambar, dan jaringan Diskriminator D, yang mencoba membedakan antara gambar yang dihasilkan 'nyata' dan 'palsu'. Satu dapat melatih jaringan ini terhadap satu sama lain dalam permainan min-max di mana generator berusaha untuk menipu diskriminator secara maksimal sementara secara bersamaan diskriminator berusaha untuk mendeteksi contoh mana yang palsu.

Di mana  $z$  adalah "kode" laten yang sering disampel dari distribusi sederhana seperti distribusi normal. GAN bersyarat adalah perpanjangan dari GAN di mana generator dan diskriminator menerima variabel pengkondisian tambahan  $c$ , menghasilkan  $G(z, c)$  dan  $D(x, c)$ . Formulasi ini memungkinkan  $G$  untuk menghasilkan gambar yang dikondisikan pada variabel  $c$ . Dibawah ini merupakan gambar formulasi dari GAN.

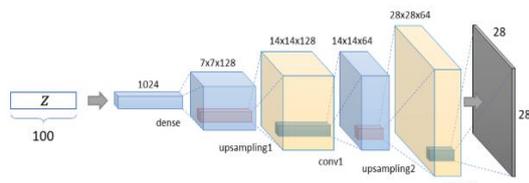
$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x) + E_{z \sim p_z(z)}] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

Gambar 1. Formulasi dari GAN

### B. Generator [3]

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.

Tugas Generator adalah dilatih untuk bisa membuat sekumpulan gambar palsu, yang saat dilihat oleh Discriminator, Discriminator tidak bisa membedakan antara asli dan palsunya. Gambar 2 adalah contoh model jaringan generator. Input Generator adalah sebuah array yang dibangkitkan secara random, disebut “seed”, pada gambar seed adalah 1x100. Generator akan mengubahnya menjadi sebuah gambar berukuran pixel NxN, pada gambar adalah 28x28 menggunakan Convolutional Neural Network.

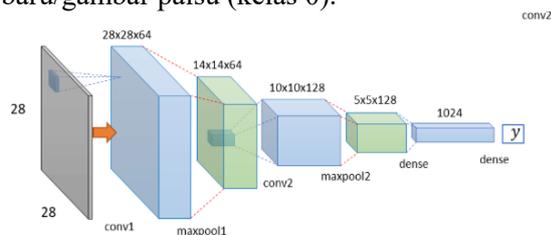


Gambar 2. Model Jaringan Generator

C. Diskriminator [3]

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. Gambar-gambar Generator yang berhasil dikenali Discriminator sebagai gambar “palsu”, akan dikembalikan sebagai feedback pada jaringan Generator.

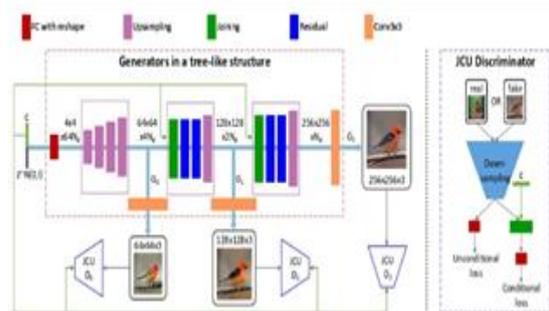
Gambar 3 adalah model jaringan Diskriminator merupakan CNN yang menerima input image berukuran [28,28] dan menghasilkan angka biner yang menyatakan apakah input gambar merupakan gambar dari dataset asli (kelas 1) atau merupakan gambar baru/gambar palsu (kelas 0).



Gambar 3. Model Jaringan Diskriminator

III. METODOLOGI

Dengan algoritme GAN bisa menggambar dengan penjelasan sederhana seperti “cat.” akan tetapi, kualitas gambarnya berantakan kalau memakai keterangan rumit seperti “a beautiful cat with blue eyes and long tail.” Penulis mencoba membuat gambar sintesis dari deskripsi teks adaptasi kode laten dari algoritme GAN [2]. Gambar 4 adalah arsitektur dari algoritme GAN



Gambar 4. Arsitektur Algoritme GAN

IV. HASIL DAN PEMBAHASAN

Pada setiap gambar yang akan dihasilkan dilatih dengan pemberian beberapa deskripsi teks untuk 1 buah gambar. Tabel 1 adalah contoh deskripsi teks untuk sebuah gambar bunga.

Tabel 1. Deskripsit Teks Sebuah Gambar Bunga

No	Deskripsi Teks
1	The flower are pink in color and have a yellow center
2	This flower is pink and white in color.
3	This flower has large pink petals and a white stigma in the center

Gambar 5 adalah gambar yang dihasilkan dari deskripsi teks tabel 1.



Gambar 5. Gambar Bunga Sintesis dari Deskripsi Teks

Gambar yang dihasilkan masih tidak detail dalam memaknai sebuah deskripsi teks. Hasil gambar dapat lebih puitis atau menginspirasi ketika dicoba menggunakan puisi, lirik, atau kutipan buku. Contoh percobaan lainnya adalah

dari deskripsi teks "Be the good, because God loves the goodness." Gambar 6 adalah hasil gambar sintesisnya.



Gambar 6. Gambar Pemandangan Sintesis dari Deskripsi Teks.

Berbagai gambar dataset publik tersedia untuk melatih algoritme GAN, hal ini dapat dilakukan untuk dipakai pada penelitian mendatang supaya hasil gambar sintesis yang dihasilkan lebih detail, memiliki makna dari maksud deskripsi teks. Misalkan dataset gambar bunga Oxford-102 yang memiliki 8.189 gambar bunga dari 102 kategori berbeda. Dataset telah dibuat dari bunga yang dipilih untuk umum. Gambar memiliki variasi skala besar, pose dan cahaya. Selain itu, ada kategori yang memiliki variasi besar di dalam kategori dan beberapa kategori yang sangat mirip. Dataset divisualisasikan menggunakan isomap dengan fitur bentuk dan warna. Setiap gambar memiliki sepuluh teks teks yang menggambarkan gambar bunga dengan cara yang berbeda. Setiap kelas terdiri dari kisaran antara 40 dan 258 gambar.

## V. PENUTUP

Penelitian ini telah berhasil menciptakan gambar dari deskripsi teks sederhana menjadi sebuah foto menggunakan algoritme GAN. Salah satu masalah paling menantang di bidang machine learning dalam melakukan sintesis gambar adalah gambar yang dihasilkan belum banyak sesuai dengan deskripsi teks, belum menghasilkan objek detail, dan algoritme yang dibuat masih ambigu dalam menghasilkan gambar apabila diberikan dataset yang lebih besar dan beragam.

## DAFTAR PUSTAKA

- [1] Tao Xu and Pengchuan Zhang and Qiuyuan Huang and Han Zhang and Zhe Gan and Xiaolei Huang and Xiaodong He, "AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks" in arXiv, 2017.
- [2] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014
- [3] F. Satya, "Menginvestigasi Kemampuan Generative Adversarial Network," Medium.com. [Online]. Available: <https://medium.com/ailab-telu/menginvestigasi-kemampuan-generative-adversarial-network-1d5d2351cbc0>. [Accessed: 06-Apr-2020].
- [4] Sarah Emerson, "Kecerdasan Buatan Sudah Bisa Menggambar Sesuai Deskripsi Kita dan Hasilnya Mengerikan". [Online]. Available: [https://www.vice.com/id\\_id/article/7xqj89/kecerdasan-buatan-sudah-bisa-menggambar-sesuai-deskripsi-kita-dan-hasilnya-mengerikan](https://www.vice.com/id_id/article/7xqj89/kecerdasan-buatan-sudah-bisa-menggambar-sesuai-deskripsi-kita-dan-hasilnya-mengerikan) [Accessed: 06-Apr-2020].