



**UNIVERSITAS BHAYANGKARA JAKARTA RAYA**  
**FAKULTAS ILMU KOMPUTER**

Kampus I: Jl. Harsono RM No. 67, Ragunan, Pasar Minggu, Jakarta Selatan 12550  
Telepon: (021) 27808121 – 27808882  
Kampus II: Jl. Raya Perjuangan, Marga Mulya, Bekasi Utara, Jawa Barat, 17142  
Telepon: (021) 88955882, Fax.: (021) 88955871  
Web: [fasilkom.ubharajaya.ac.id](http://fasilkom.ubharajaya.ac.id), E-mail: [fasilkom@ubharajaya.ac.id](mailto:fasilkom@ubharajaya.ac.id)

**SURAT TUGAS**

Nomor: ST/1313/I/2023/FASILKOM-UBJ

1. Dasar: Kalender Akademik Universitas Bhayangkara Jakarta Raya Tahun Akademik 2022/2023.
2. Dalam rangka mewujudkan Tri Dharma Perguruan Tinggi untuk Dosen di Universitas Bhayangkara Jakarta Raya maka dihimbau untuk melakukan Penelitian.
3. Sehubungan dengan hal tersebut di atas, maka Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya menugaskan:

NO.	NAMA	NIDN	JABATAN	KETERANGAN
1.	Dr. Rakhmi Khalida, S.T., M.M.S.I.	0304099201	Dosen Tetap Prodi Informatika	Sebagai Penulis Pertama

Membuat Artikel Ilmiah dengan judul "**Concatenate Word Embedding for Text to Image through Generative Adversarial Network**" pada *media Institute of Electrical and Electronics Engineers (IEEE) Xplore*, 20 Januari 2023, DOI: 10.1109/ICIMCIS56303.2022.10017727, dengan tanggal Konferensi: 16 s.d 17 November 2022 di *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*.

4. Demikian penugasan ini agar dapat dilaksanakan dengan penuh rasa tanggung jawab.

Jakarta, 09 Januari 2023  
**DEKAN FAKULTAS ILMU KOMPUTER**



**Dr. Dra. Tyastuti Sri Lestari, M.M.**  
NIP. 1408206



**FAKULTAS ILMU KOMPUTER  
UPN VETERAN JAKARTA**

# PROCEEDINGS



**4<sup>th</sup> INTERNATIONAL CONFERENCE**

# ICIMCIS

**INFORMATICS, MULTIMEDIA, CYBER, AND INFORMATION SYSTEM**

ICIMCIS PROCEEDINGS 2022

**2022**

Technical Co-Sponsored by  
  
INDONESIA SECTION

# **2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)**

## **Organizing Committee International Advisory Committee**

- Dr. Ing Wahyudi Hasbi, S.Si, M.Kom. - IEEE IS Chair  
Prof. Dr. Ir. Gamantyo Hendranto, Ph.D. - IEEE IS Vice-Chair  
Ir. Linawati, M.Eng.Sc., Ph.D. - IEEE IS Secretary  
Dr. Karlisa Priandana, M.Eng - IEEE IS Vice Secretary  
Dr. Kemas Muslim Lhaksmana - IEEE IS Treasurer  
Dr. Prima Dewi Purnamasari, S.T., M.Sc - IEEE IS Vice Treasurer  
Dr. Anter Venus, MA, Comm – Rector Universitas Pembangunan Nasional Veteran Jakarta, Indonesia  
Dr. R. Dudy Heryadi, M.Si – Vice Rector for Academic Affairs Universitas Pembangunan Nasional Veteran Jakarta, Indonesia  
Dr. Ec. Prasetyo Hadi, SE., MM., CFMP – Vice Rector for General and Financial Affairs Universitas Pembangunan Nasional Veteran Jakarta, Indonesia  
Dr. dr. Ria Maria Theresa, Sp.KJ., MH. – Vice Rector for Student Affairs and Cooperation Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

## **Steering Committee**

- Prof. Teddy Mantoro – Universitas Sampoerna, Indonesia  
Prof. Achmad Nizar Hidayanto - Universitas Indonesia, Indonesia  
Prof. Dr. Achmad Benny Mutiara Q.N. – Universitas Gunadarma, Indonesia  
Prof. Dra. Sri Hartati, M.Sc., Ph.D – Universitas Gajah Mada, Indonesia  
Prof. Dr.Ir. Agus Buono, M.Si, M.Kom – IPB University, Indonesia  
Prof. Dr. Ir. Siti Nurmaini, MT – Universitas Sriwijaya, Indonesia  
Dr. Djuniadi, MT – Universitas Negeri Semarang, Indonesia  
Dr. Ermatita, M.Kom – Universitas Sriwijaya, Indonesia  
Dr. Didit Widiyanto – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia  
Dr. Tjahjanto – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia  
Dr. Widya Cholil, S.Kom., M.IT – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia  
Dr. Bambang Saras – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia  
Dr. Deris Setiawan - Universitas Sriwijaya, Indonesia

## **General Chair**

- Dr. Widya Cholil, S.Kom., M.IT – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

## **General Co-Chairs**

- M. Octaviano Pratama, S.Kom, M.Kom. – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

### **Secretarial**

Kraugusteeliana, M.Kom., MM – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Diyah Retnowati, S.Kom – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Saimin, S.Kom – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

### **Event, Publicity, Website & Documentation Division**

Ati Zaidiah, S.Kom., MTI – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Andhika Octa Indarso, MMSI – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Astrianto Afandi, Amd.Kom – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Didit Suryahartono, S.Kom – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Mochammad Fariz Setyawan, S.Kom – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

### **Proceeding Division**

Erly Krisnanik, S.Kom, MM – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Artika Arista, S.Kom., MMSI – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Ika Nurlaili Isnainiyah, S.Kom, M.Sc – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Helena Nurramdhani Irmada, S.Pd., M.Kom. – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Muhammad Adrezo, S.Kom, M.Sc – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Ria Astriratma, S.Kom, M.Cs - Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Neny Rosmawarni, M.Kom - Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Muhammad Panji Muslim, S.Pd., M.Kom. - Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Mustofa Galih Pradana, M.Kom - Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Triando Damiri Burlian, S.Kom, M.S.Eng, M.Sc. – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Novi Trisman Hadi, S.Pd., M.Kom. - Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

### **Financial**

Rudhy Ho Purabaya, SE., MMSI – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Yami - Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Suryadi, Amd.Kom - Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

### **Administration & Logistic**

Musenah, S.Pd, M.Kes – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Zaenudin – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

Abdul Rahman – Universitas Pembangunan Nasional Veteran Jakarta, Indonesia



## Table of Content

### **IT Capabilities Analysis Towards the Organization Performance in the Virtual Organization: A Case Study of ABC Multifinance Company in Indonesia..... 1**

Dewi Rokhmah Pyriana (University of Indonesia, Indonesia); Hidayat Akbar (Universitas Indonesia, Indonesia); Muhammad Muslim Rifa'i (University of Indonesia, Indonesia); Sandriyana Kusuma Wardani and Suci Inayah (Universitas Indonesia, Indonesia); Muhammad Rifki Shihab (Faculty of Computer Science, Universitas Indonesia, Indonesia) ..... 1

### **E-Warung Systems Development ..... 7**

Cornelia Tri Wahyuni (Universitas Bina Darma, Indonesia); Darius Antoni (Universitas Indo Global Mandiri, Indonesia); Widya WH Cholil, MIT (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia) ..... 7

### **Diagnosing the Impact of Digital Resources and Remote Leadership on Telework Productivity ..... 12**

Nopriadi Saputra (Bina Nusantara University, Indonesia); Kiki Suidiana (Telkom University, Indonesia) ..... 12

### **Public Government Web Based Services Quality Assesment Using Webqual and Govqual in Palembang City Government..... 19**

Widya WH Cholil, MIT (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Lalilatur Rahmi (Politeknik Negeri Sriwijaya, Indonesia); Anita Muliawati, Jayanta Jayanta, Rio Wirawan and Helena Nurramdhani Irmanda (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia) ..... 19

### **Empirical Review of M-Banking User Satisfaction Using End User Computing Satisfaction (EUCS) ..... 25**

Dedy Syamsuar (Bina Nusantara University, Indonesia); Widya WH Cholil, MIT (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Rio Ramadhan (Universitas Bina Darma, Indonesia); Ruth Mariana Bunga Wadu (Jalan RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Indonesia); Ati Zaidiah (UPN Veteran Jakarta, Indonesia); Nia Oktaviani (Universitas Bina Darma, Indonesia)..... 25

### **Small Medium Enterprise Social Media Ranking Using Analytic Network Process ..30**

Khairun Nisa Meiah Ngafidin, Citra Wiguna and Rona Nisa Sofia Amriza (Institut Teknologi Telkom Purwokerto, Indonesia) ..... 30

### **Identifying Shill Bidder Presence in Online Auction Using Dynamic Detection Strategy ..... 36**



Felesia Felesia, Ivanco Gratio Hartono, Monica Ruth Fabiola Winata, Henry Lucky and Andry Chowanda (Bina Nusantara University, Indonesia)..... 36

**Development of Mobile-Based Information Letter of Rukun Tetangga and Rukun Warga With Mern Stack Method..... 43**

Gregorius Wisnu Wijanarko Tiba and Haris Rangkuti, Mr (Bina Nusantara University, Indonesia) ..... 43

**Development of E-Learning Applications to Support Student Learning..... 49**

Arvin Rizki Andriaputra, Dicky Bima Pradana Resmana, Nico Marzuki and Haris Rangkuti, Mr (Bina Nusantara University, Indonesia) ..... 49

**Use of Blockchain for Designing Digital Documents in Public Services..... 55**

Darius Antoni (Universitas Indo Global Mandiri, Indonesia); Muhammad Izman Herdiansyah (Universitas Bina Darma, Indonesia); Muhamad Akbar (Universitas Sriwijaya, Indonesia & Universitas Bina Insan, Indonesia); Widya WH Cholil, MIT (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Hadi Syaputra (Universitas Sriwijaya & Universitas Bina Darma, Indonesia) ..... 55

**The Role of Vlog Content and Vlog Enjoyment on Developing Marketable Travel E-Commerce to Millennial Audiences ..... 61**

Martinus Fieser Sitinjak (Bina Nusantara University, Indonesia); Azizah Omar (USM, Malaysia); Yopy Maulana (Bina Nusantara University, Indonesia) ..... 61

**A Systematic Review of Business Process Management in E-Commerce ..... 66**

Thalea Christy Nathaniela, Agung Purnomo and Satria Fadil Persada (Bina Nusantara University, Indonesia); Elsa Rosyidah (Universitas Nahdlatul Ulama Sidoarjo, Indonesia) ..... 66

**Effectiveness Analysis of Using the Vehicle Tools Feature on ERP Systems With Delone & McLean Approach at PT. Putra Mulia Telecommunication..... 71**

Azizah Tsana Putri, Haffiz Galuh Pratama Pratama, Orchidea Safira and Sucianna Ghadati Rabiha (Bina Nusantara University, Indonesia) ..... 71

**Applying for E-Signature Approval With TOGAF Framework to Improve Productivity: Case Study SAP Document Management System..... 77**

Cynthia Dewi Tanuraharja, Kania Alma Tiara, Gunawan Wang and Hendra Alianto (Bina Nusantara University, Indonesia) ..... 77

**Analysis of Stakeholder Satisfaction Using ITILv4 for Data Management in Telkomsat ERP..... 82**

Nugroho Wibisono (Swiss German University & PT Telkom Satelit Indonesia (Telkomsat), Indonesia); Heru Ipung, Amin Soetomo and Eka Budiarto (Swiss German



University, Indonesia); Ayu Putri Sakinah (IPB University, Indonesia & PT Telkom Satelit Indonesia, Indonesia); Rian Fitriana (PT Telkom Satelit Indonesia, Indonesia)82

**A Literature Review Knowledge Management for Traditional Retail Business Strategies..... 88**

Dien Novita (Universitas Sriwijaya & Universitas MDP, Indonesia); Ermatita Ermatita (Sriwijaya University & Computer Science Faculty, Indonesia); Samsuryadi Samsuryadi and Dian Palupi Rini (Universitas Sriwijaya, Indonesia)..... 88

**Financial Auditor Efficiency and Effectiveness in Remote Audit Transition Due to Pandemic of Covid-19..... 94**

Bambang Leo Handoko and Ang Swat Lin Lindawati (Bina Nusantara University, Indonesia); Hery Harjono Muljo (School of Information Systems, Bina Nusantara University, Indonesia) ..... 94

**Functional Requirements and Traceability System Information Modeling on Java Preanger Coffee Supply Chain ..... 99**

Bonang W Ligar (Gunadarma University, Indonesia); Sarifuddin Madenda (Gundarma University, Indonesia); Sutrisno S Mardjan (IPB University, Indonesia); Tubagus Maulana Kusuma (Gunadarma University, Indonesia)..... 99

**Analysis and Design of Supply Chain Management System at CV Putra Nusantara Mandiri ..... 105**

Nomane Danke Zaidane, Sheline Rizkyka Giovani, Rizki Putra Bahari and Sucianna Ghadati Rabiha (Bina Nusantara University, Indonesia) ..... 105

**Identifying Value at Risk in the Real Estate Sector Through the Application of Monte Carlo Simulation ..... 111**

Dicky Hida Syahchari and Anggraeni Woro Hapsari (BINUS University, Indonesia)111

**Model for Detect Hand Sign Language Using Deep Convolutional Neural Network for the Speech/Hearing Impaired ..... 118**

Alfredo Lorentiar Santonanda, Fransisca Cicilia Margono, Rachmat Kurniawan, Henry Lucky and Andry Chowanda (Bina Nusantara University, Indonesia)..... 118

**A Preliminary Study of Decision Support Model of Photovoltaic for Village Area in South of Sumatera..... 124**

Feby Ardianto (Universitas Muhammadiyah Palembang, Indonesia); Ermatita Ermatita (Sriwijaya University & Computer Science Faculty, Indonesia); Armin Sofijan, AS (Sriwijaya University & Faculty of Engineering, Indonesia); Vina Ayumi (Universitas Mercu Buana Jakarta, Indonesia); Handrie Noprisson (Universitas Mercu Buana, Indonesia); Mariana Purba (Universitas Sjakhyakirti, Indonesia) ..... 124

**Identification of Meteorological Parameters Affecting Dengue Hemorrhagic Fever Incidences..... 130**



Faisal Asadi, Joko Pebrianto Trinugroho, Alam Ahmad Hidayat, Reza Rahutomo and Bens Pardamean (Bina Nusantara University, Indonesia) ..... 130

**Scientometric Analysis of Artificial Intelligence Research in Agriculture..... 136**

Fairuz Iqbal Maulana and Andi Pramono (Bina Nusantara University, Indonesia); Miftahul Hamim (Language Center, Bina Nusantara University, Indonesia); Sukeipah Yuli Prihatin (Politeknik Kota Malang, Indonesia); Rahman Arifuddin (University of Merdeka Malang, Indonesia)..... 136

**Indonesian Traditional Food Image Recognition Using Convolutional Neural Network ..... 142**

Gredion Prajena, Jeklin Harefa, Alexander Alexander, Bernhard Owen Josephus and Ardianto Hermawan Nawir (Bina Nusantara University, Indonesia) ..... 142

**Calculating the Similarity of Indonesian Sentences Using Latent Semantic Indexing Based on KBBI ..... 148**

Muhammad Panji Muslim (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Arief Fatchul Huda (Universitas Islam Negeri Bandung & Telkom University, Indonesia) ..... 148

**Evaluation of Software Quality Assurance Silampari Smart City of Lubuklinggau Based on ISO/IEC 25010:2011 Analysis Model ..... 154**

Yayan Indra Irawan (Bina Darma University, Indonesia); Edi Surya Negara (Universitas Bina Darma, Indonesia)..... 154

**Time-Series Analysis of Correlation Between Climatic Parameters and Dengue Fever in Indonesia..... 161**

Rudi Nirwantono (Binus University, Indonesia); Joko Pebrianto Trinugroho, Digdo Sudigyo, Alam Ahmad Hidayat and Bens Pardamean (Bina Nusantara University, Indonesia) ..... 161

**A Fatigue Detection Model Based on Convolutional Neural Network ..... 166**

Daniel Wijaya, Michael Stanley, Pascal Wilman, Henry Lucky and Andry Chowanda (Bina Nusantara University, Indonesia) ..... 166

**A Literature Review on the Impact of Effective Management in Cyber Security System Performance ..... 172**

Luis F Salim, Stephanus Harjono, Ferdinand Gunawan, Jurike Moniaga and Indra D Rianto (Bina Nusantara University, Indonesia) ..... 172

**Comparative Analysis of Usability, Performance, and Security of Open-Source, Windows-Based Password Manager Applications Based on ISO/IEC 25010..... 178**

Lettisia Nurdayenti (Badan Siber dan Sandi Negara, Indonesia); Amiruddin Amiruddin (Politeknik Siber dan Sandi Negara & Badan Siber Dan Sandi Negara, Indonesia) .. 178

**A Case Study of Digital Forensic Readiness Level Measurement Using DiFRI Model 184**

Ilham Zulfahmi Baiquni (Badan Siber dan Sandi Negara, Indonesia); Amiruddin Amiruddin (Politeknik Siber dan Sandi Negara & Badan Siber Dan Sandi Negara, Indonesia) ..... 184

**Cyber Threats and Scams in FinTech Organizations: A Brief Overview of Financial Fraud Cases, Future Challenges, and Recommended Solutions in Bangladesh ..... 190**

Md. Jafrin Hossain, Rejuan Haque Rifat and Mahadi H Mugdho (BRAC University, Bangladesh); Mohona Jahan (Mohakhali, Dhaka, Bangladesh); Annajiat Alim Rasel (Brac University, Bangladesh); Muhammad Abdur Rahman (BRAC University, Bangladesh)..... 190

**Impact Analysis of Crypto Miner Malware Attacks Using Android Debug Bridge (ADB) Vulnerabilities via TCP/IP on Android-Based Raspberry Pi 4 IoT Device ... 196**

Aidil Yusuf Priadi (National Cyber and Crypto Polytechnic, Indonesia); Arizal Arizal (Politeknik Siber dan Sandi Negara, Indonesia) ..... 196

**Flubot Malware Hybrid Analysis on Android Operating System..... 202**

Hanifah Salsabila (Poltek Siber dan Sandi Negara, Indonesia); Syafira Mardhiyah (Politeknik Siber Dan Sandi Negara, Indonesia); Raden Budiarto Hadiprakoso (Poltek Siber dan Sandi Negara, Indonesia) ..... 202

**Implementation of Rail Fence Cipher and Myszkowski Algorithms and Secure Hash Algorithm (SHA-256) for Security and Detecting Digital Image Originality ..... 207**

Rojali Rojali, Zulfany Rasjid and Justin Cliff Matthew (Bina Nusantara University, Indonesia) ..... 207

**Implementation of PRESENT Algorithm on Contactless Access Control Using Raspberry Pi..... 213**

Nazela Khairani Putri and Dion Ogi (Politeknik Siber dan Sandi Negara, Indonesia)213

**Comparison of IPv6 Dynamic Routing Protocols on Routing Hole Handling ..... 219**

Novi Trisman Hadi (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia)219

**Comparison of Machine Learning Techniques on Snort for Predicting Realtime DoS and Probe Attack ..... 224**

Achmad Husein Noor Faizi, Dimas Febriyan Priambodo and Fika Dwi Rahmawati (Politeknik Siber dan Sandi Negara, Indonesia) ..... 224

**Waterfall Model for Design and Development Coffee Shop Website at Malang..... 230**

Andreas Yan Ardhiansyah, Dion Lamilga Sudiono Putra, Jsven Septian K, Nadya Budhianto and Fairuz Iqbal Maulana (Bina Nusantara University, Indonesia) ..... 230

**Low Fidelity Prototype Design to Facilitate BEELIC Website Navigation Process for Information Center ..... 235**

Hervina Aprilia (Universitas Bina Nusantara, Indonesia); Arsa Widitiarsa Utoyo (Bina Nusantara University & Bina Nusantara University - School of Design, Indonesia); Tobias Warbung (BINUS University & Towa Design, Indonesia); Titik Endahyani (Queensland University of Technology, Indonesia); Axel Juan Sukmawan (Bina Nusantara University, Indonesia) ..... 235

**Hybrid Protege Training Program in Corporate Setting: An Evaluative Case Study 239**

Nurul Hijja Mazlan (Universiti Teknologi MARA, Malaysia); Nor Zanariah Talib (INSTITUT AMINUDDIN BAKI, Malaysia); Nur Aqilah R M Kamal (Universiti Teknologi MARA, Malaysia) ..... 239

**A User-Centered Design of Natural Language Processing for Maternal Monitoring Chatbot System ..... 244**

Sandra Hakiem Afrizal (Universitas Binawan, Indonesia); Nashrul Hakiem (Universitas Islam Negeri Syarif Hidayatullah, Indonesia); Adhistya Erna Permanasari (Universitas Gadjah Mada, Indonesia); Hadid Albab (UIN Syarif Hidayatullah, Indonesia); Guardian Yoki Sanjaya and Lutfan Lazuardi (Universitas Gadjah Mada, Indonesia)..... 244

**Manufacturing Company Innovative Level Assessment Case Study of Volta Indonesia Semesta Incorporated ..... 249**

Bambang Purwanggono and Ratna Purwaningsih (Diponegoro University, Indonesia); Arfan Bakhtiar (Technical University of Berlin-Germany & Diponegoro University-Indonesia, Germany); Febrina Agusti (Duta Bangsa University, Indonesia); Kumara Pinasthika Dharaka (University of Bina Nusantara, Indonesia)..... 249

**User Interface Design of Seminar and Workshop Information Mobile Apps for Self-Development ..... 254**

Jeremy Nathanael Wijaya and Yudhistya Ayu Kusumawati (Bina Nusantara University, Indonesia) ..... 254

**Concatenate Word Embedding for Text to Image Through Generative Adversarial Network..... 259**

Rakhmi Khalida, rkhalida (Universitas Gunadarma, Indonesia); Sarifuddin Madenda (Gundarma University, Indonesia); Suryadi Harmanto (Gunadarma University, Indonesia); I Made Wiryana (Gundarma University, Indonesia)..... 259

**Analysis of SMEs Consideration in Adopting New Technology Using Technology Acceptance Model ..... 265**

Iston Utama, Mulyani Karmagatri, Dian Kurnianingrum and Okky Rizkia Yustian (Bina Nusantara University, Indonesia) ..... 265

**The Usability Analysis of Human Resource Information System (HRIS) Using Usability Scale System and Concurrent Think Aloud..... 270**

Ismayana Teguh Pratama, Rahmatika Putri and Rian Fernanda (Universitas Bina Nusantara, Indonesia); Sunardi Action (Bina Nusantara University, Indonesia)..... 270

**Comparison of Machine Learning Methods in Sentiment Analysis PeduliLindungi Applications..... 276**

Widya WH Cholil, MIT (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Febriyanti Panjaitan (Universitas Sriwijaya & Universitas Bina Darma, Indonesia); Ferdiansyah Ferdiansyah (Universitas Bina Darma & Universiti Teknologi Malaysia, Indonesia); Artika Arista, Ria Astriratma and Tri Rahayu (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia) ..... 276

**Performance Comparison of Rabin-Karp Algorithm and Winnowing Algorithm for Document Abstraction Similarity Detection ..... 281**

Anggit Dwi Hartanto and Yoga Pristyanto (Universitas Amikom Yogyakarta, Indonesia); Andy Saputra (Amikom University Yogyakarta, Indonesia); Eli Pujastuti, Atik Nurmasani and Ika Asti Astuti (Universitas Amikom Yogyakarta, Indonesia).. 281

**Improved Chest X-Ray Image Quality Using Median and Gaussian Filter Methods 287**

Lucky Indra Kesuma (Universitas Sriwijaya & Universitas Sjakhyakirti Palembang, Indonesia); Ermatita Ermatita (Sriwijaya University & Computer Science Faculty, Indonesia); Erwin E and Purwita Sari (Universitas Sriwijaya, Indonesia); Rudhy Purabaya (UPN Veteran Jakarta, Indonesia)..... 287

**Implementation of Reinforcement Learning in 2D Based Games Using Open AI Gym..... 293**

Bayu Setiaji (Universitas AMIKOM Yogyakarta, Indonesia); Eli Pujastuti and Muhammad Fairul Filza (Universitas Amikom Yogyakarta, Indonesia); Ahlihi Masruro (Universitas AMIKOM Yogyakarta, Indonesia); Yusuf Ashidicki Pradana (Universitas Amikom Yogyakarta, Indonesia) ..... 293

**Improved Support Vector Machine (SVM) Performance on Go-Jek Service Review Classification Using Particle Swarm Optimization (PSO)..... 298**

Windha MP Duhita (Universitas AMIKOM Yogyakarta, Indonesia); Haryoko Haryoko (University of AMIKOM Yogyakarta, Indonesia)..... 298

**Machine Learning Model for Detecting Fake News Content in Indonesian-Language Online Media ..... 302**

Inggrid Yanuar Risca Pratiwi (Politeknik Masamy Internasional, Indonesia); Anggit Ferdita Nugraha (Universitas AMIKOM Yogyakarta, Indonesia); Yoga Pristyanto, Rifda Faticha Alfa Aziza and Jeki Kuswanto (Universitas Amikom Yogyakarta, Indonesia); Ibnu Hadi Purwanto (Universitas AMIKOM Yogyakarta, Indonesia) .... 302

**Problem Identification on Cardiovascular Disease Prevention Using Artificial Intelligence: A Literature Review ..... 308**

Boby Siswanto (Bina Nusantara University, Indonesia)..... 308

**UI Pattern Analysis in Obtaining Highest Success Rate in University Admission Website..... 314**

Arroyan Cahyaning Ratri and Eli Pujastuti (Universitas Amikom Yogyakarta, Indonesia); Muh Hanafi (Universitas Amikom Yogyakarta, Indonesia & Time Excellindo, Malaysia); Nuri Cahyono (Universitas Amikom Yogyakarta, Indonesia)314

**Teman Berlingdung: Mobile-Based Student Counseling Service Application ..... 320**

Muhammad Fahrizal, Muhammad Raehan M Safitroh, Hanifa W Damayanti and Ihrom Wahyuni (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Chaerul Ilmi Al Ahyari (JL RS. FATMAWATI & UPN Veteran Jakarta, Indonesia); Ria Astriratma (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia)..... 320

**BERT-Based Topic Modeling Approach for Malaria Research Publication..... 326**

Alam Ahmad Hidayat (Bina Nusantara University, Indonesia); Rudi Nirwantono (Binus University, Indonesia); Arif Budiarto and Bens Pardamean (Bina Nusantara University, Indonesia) ..... 326

**A Video Filtering Method Using Greyscale and Canny Edge Detector in LabVIEW332**

Putri Wulandari (University of Al Azhar Indonesia, Indonesia); Dwi Astharini (Universitas Al Azhar Indonesia, Indonesia); Wahyu Caesarendra (Universiti Brunei Darussalam & Diponegoro University, Brunei Darussalam); Triwiyanto Triwiyanto (Health Polytechnic Ministry of Health Surabaya, Indonesia); Maciej Sulowicz (Cracow University of Technology, Poland); Tjahjanto Tjahjanto (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia) ..... 332

**Comparison Analysis of Data Sending Performance Using the Cayenne and ThingSpeak IoT Platform ..... 337**

Kirana Astari Pranoto (Universiti Brunei Darussalam, Brunei Darussalam); Yogi Reza Ramadhan (CV. Rekeyasa Desain Manufaktur, Indonesia); Wahyu Caesarendra (Universiti Brunei Darussalam & Diponegoro University, Brunei Darussalam); Triwiyanto Triwiyanto (Health Polytechnic Ministry of Health Surabaya, Indonesia); Adam Glowacz (AGH University of Science and Technology, Poland); Santanu Kumar Dash (Vellore Institute of Technology, India); Bambang T Wahyono (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia) ..... 337

**Naive Bayes Classification Framework Model for Optimizing Prediction of Agrotourism Products Orange Gerga..... 343**

Yogi Isro Mukti, YIM (Institut Teknologi Pagar Alam & Teknik Informatika, Indonesia); Vike Itteridi and Iskandar Sulaini (Institut Teknologi Pagar Alam,



Indonesia); Yuni Widiastiwi (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia) ..... 343

**Comparison ADAM-Optimizer and SGDM for Classification Images of Rice Leaf Disease..... 348**

Pujo Hari Saputro (Sam Ratulangi University, Indonesia); Dhina Puspasari Wijaya Action (Alma Ata University, Indonesia); Musthofa Galih Pradana (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Dyah Listianing Tyas (Universitas Trinita Manado, Indonesia); Wahyuni Fithratul Zalmi (Universitas Sam Ratulangi, Indonesia) ..... 348

**Ensemble Learning on Sentinel-1A Imagery for Garlic Field Classification..... 354**

Imas Sukaesih Sitanggung (Bogor Agricultural University, Indonesia); Muhammad Asyhar Agmalaro, Ahmad Al Banjaran Coeur d'Alene and Annisa Annisa (IPB University, Indonesia) ..... 354

**Sperm Cell Classification System Carrying X or Y Chromosome in Human With CNN Algorithm..... 360**

I Gede Susrama Mas Diyasa (Universitas Pembangunan Nasional Veteran Jawa Timur, Indonesia); Kraugusteeliana Tambunan (Universitas Pembangunan Nasional, Indonesia); Wahyu S J Saputra (University of Pembangunan Nasional Veteran Jawa Timur, Indonesia) ..... 360

**Detection of Ovarian Cancer Risk Level Using the Web-Based Dempster Shafer Method ..... 365**

Kraugusteeliana Tambunan (Universitas Pembangunan Nasional, Indonesia); Nursaka Putra (STIKES Al-Insyirah Pekanbaru, Indonesia); Miandhani Denniz Yuniar (Universitas Selamat Sri, Indonesia); Dasril Aldo (Institut Teknologi Telkom Purwokerto, Indonesia); Anik Rahmawati and Viro Dharma Saputra (Universitas Selamat Sri, Indonesia) ..... 365

**English-To-Sundanese Translation Using Neural Machine Translation: Collection and Analytics..... 370**

M. Octaviano Pratama (UPN Veteran Jakarta, Indonesia); Pamela Kareen (BISA AI Academy, Indonesia); Ermatita Ermatita (Sriwijaya University & Computer Science Faculty, Indonesia); Peng Choo and Miroslav Kosteki (STEMSEL Australia, Australia); Desmond Devendran (Logical Operation, USA) ..... 370

**The Effect of Imbalanced Classes on Retinal Blood Vessel Segmentation..... 375**

Ahmad Fauzi and Raju Wandira (UIN Imam Bonjol, Indonesia); Kraugusteeliana Tambunan (Universitas Pembangunan Nasional, Indonesia) ..... 375

**Comparison the New Computer Sciences Study Programs in Indonesia Using PyTrends..... 379**



Tora Fahrudin (Telkom University, Indonesia); Asniar Asniar (Telkom University);  
 Muhammad Faizul Ula (Telkom University, Indonesia) ..... 379

**Fine-Tuned MobileNetV2 and VGG16 Algorithm for Fish Image Classification ..... 384**

Devi Fitriana, Kristien Suryaningrum and Noviyanti Tri Mareta Sagala (Bina Nusantara University, Indonesia); Vina Ayumi (Universitas Mercu Buana Jakarta, Indonesia); Siew Mooi Lim (Tunku Abdul Rahman University College, Malaysia) . 384

**Development of Educational Application About Mental Disorders With Augmented Reality for Android..... 390**

Jonathan Frederik Maringka, Rahadian Hargyo Pambuko and Mochammad Haldi Widiyanto (Bina Nusantara University, Indonesia) ..... 390

**Implementation of Feature Selection Based on Particle Swarm Optimization and Genetic Algorithm on Support Vector Regression Algorithm to Predict Student Performance ..... 395**

M. Riki Apriyadi (Doctoral Student of Engineering Science, Faculty of Engineering University of Sriwijaya Palembang, Indonesia); Ermatita Ermatita (Sriwijaya University & Computer Science Faculty, Indonesia); Dian Palupi Rini (Sriwijaya University & Computer Science Faculty, Indonesia); Yulnelly Yulnelly (UPN Veteran Jakarta, Indonesia) ..... 395

**Maturity Classification of Golek Mango Using Fuzzy C Means Clustering Method Based on HSI and YCbCr Color Space Transformation ..... 401**

Ibrohim Habiburrohman (Universitas Sebelas Maret, Indonesia); Esti Suryani (University of Sebelas Maret, Indonesia); Wiharto Wiharto (Universitas Sebelas Maret, Indonesia) ..... 401

**Implementation of Convolutional Neural Network Algorithm for Android-Based Classification of Taekwondo Martial Arts ..... 407**

Fauzan Akmal Mahdi and Desta Sandya Prasvita (UPN Veteran Jakarta, Indonesia); Theresiawati Theresiawati (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia) ..... 407

**Chatbot Application to Automate Services in FnB Business Using Seq2Seq LSTM 413**

Ethan Wang and Winata Liadylova Putera (BINUS University, Indonesia); Henry Lucky and Andry Chowanda (Bina Nusantara University, Indonesia) ..... 413

**Analysis Performance of Support Vector Regression (SVR) for Congestion Prediction..... 418**

Nindy Irzavika (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Putri Saptawati (Bandung Institute of Technology, Indonesia) ..... 418

**A Combination of Image Enhancement and U-Net Architecture for Segmentation in Identifying Brain Tumors on CT-SCAN Images ..... 423**

Anita Desiani (Universitas Sriwijaya, Indonesia); Muhammad Adrezo (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Nyayu Chika Marselina (Universitas Sriwijaya, Indonesia); Muhammad Arhami (Politeknik Negeri Lhokseumawe, Indonesia); Aulia Salsabila and Muhammad Gibran Al Filambany (Universitas Sriwijaya, Indonesia) ..... 423

**Implementation of Big Data Analytics for Machine Learning Model Using Hadoop and Spark Environment on Resizing Iris Dataset..... 429**

Tresna Maulana Fahrudin, Prismahardi Aji Riyantoko and Kartika Maulida Hindrayani (Universitas Pembangunan Nasional Veteran Jawa Timur, Indonesia) ..... 429

**Comparison of Kernel Support Vector Machines in Conducting Sentiment Analysis Review of Buying Chips on the Shopee E-Marketplace in Indonesian..... 435**

Muhammad Eka Purbaya, Dioviando Putra Rakhmadani, Maliana Puspa Arum and Luthfi Zian Nasifah (Institut Teknologi Telkom Purwokerto, Indonesia) ..... 435

**Application of Median Filter Method for Classification of Oil Palm Tree on LiDAR Images ..... 441**

Sabila Rafani Aliandra and Desta Sandya Prasvita (UPN Veteran Jakarta, Indonesia)441

**Application of Genetic Algorithm K-Means Clustering of Villagers Characteristics for Smart Economy ..... 445**

Eneng Tita Tosida (IPB University & Universitas Pakuan, Indonesia); Indra Permana Solihin (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Fajar Delli Wihartiko (IPB University & Universitas Pakuan, Indonesia); Farhan Naufal (Pakuan University, Indonesia) ..... 445

**Implementation of Multiple Discriminant Analysis (MDA) for Clustering Smart Village in West Java Based Podes (Potensi Desa) Database ..... 451**

Eneng Tita Tosida (IPB University & Universitas Pakuan, Indonesia); Indra Permana Solihin (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Roni Jayawinangun and Deden Ardiansyah (Pakuan University, Indonesia) ..... 451

**Clustering of Tidal Swamp Soil Based on NPK Macronutrients Using the FCM Method ..... 457**

Suryati Suryati (Universitas Sriwijaya, Indonesia & Universitas Indo Global Mandiri, Indonesia); Ermatita Ermatita (Sriwijaya University & Computer Science Faculty, Indonesia); Dian Palupi Rini and Dedik Budianta (Universitas Sriwijaya, Indonesia)457

**Comparison of Multi Layer Perceptron, Random Forest & Logistic Regression on Students Performance Test ..... 462**

Musthofa Galih Pradana (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Kenneth Palilingan (Sam Ratulangi University, Indonesia); Yuri Vanli Akay

(University of Sam Ratulangi, Indonesia); Dhina Puspasari Wijaya Action (Alma Ata University, Indonesia); Pujo Hari Saputro (Sam Ratulangi University, Indonesia) ... 462

**Investment Feasibility Analysis of Base Transceiver Station (BTS) Tower Using Cost-Benefit Analysis on PT. XYZ ..... 467**

Arvy Muhammad Muhammad Reyhan and Sarah Shafira (Institute Technology of Sepuluh Nopember, Indonesia); Apol Pribadi Subriadi (Institut Teknologi Sepuluh Nopember, Indonesia) ..... 467

**Public Services Point Network Design for Local Government ..... 473**

Darius Antoni (Universitas Indo Global Mandiri, Indonesia); Muhammad Izman Herdiansyah (Universitas Bina Darma, Indonesia); Muhamad Akbar (Universitas Sriwijaya, Indonesia & Universitas Bina Insan, Indonesia); Muhamad Taufiq (Universitas Bina Darma, Indonesia); Bayu Hananto, Didit Widiyanto and I Wayan Widi Pradnyana (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia)..... 473

**Information System Management of Palm Agriculture Using Laravel Framework 478**

Muhammad Ilham Alhari, Muharman Lubis and Faisal Budiman (Telkom University, Indonesia) ..... 478

**Investigation Learning Model for Physics Courses Using Web Application With Case-Based Approach ..... 484**

Anindita Mardiani, Erly Krisnanik, Bayu Wibisono and Bambang Saras Yulistiawan (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia)..... 484

**Knowledge Management Model Review in Relation to Marketing and Branding.... 490**

Fransiska Prihatini Sihotang (Sriwijaya University & Universitas MDP, Indonesia); Ermatita Ermatita (Sriwijaya University & Computer Science Faculty, Indonesia); Samsuryadi Samsuryadi and Dian Palupi Rini (Universitas Sriwijaya, Indonesia).... 490

**The Determinant Factors of In-App Purchase Intention of International Genshin Impact Player Mediated by Perceived Value and Game Loyalty ..... 496**

Wiwin Lestari (Universitas Bina Nusantara, Indonesia); Sofia Ranti, Nadaffa Aisy Nozwar and Arta Moro Sundjaja (Bina Nusantara University, Indonesia)..... 496

**Implementation of Haversine Algorithm and Geolocation for Travel Recommendations on Smart Applications for Backpackers in Bali ..... 504**

I Gede Susrama Mas Diyasa, Dwi Arman Prasetya, Mohammad Idhom and Anggraini Puspita Sari (Universitas Pembangunan Nasional Veteran Jawa Timur, Indonesia); Anuar Bin Mohamed Kassim (Universiti Teknikal Malaysia Melaka, Malaysia)..... 504

**The Positive Impact of ERP Usage on Business Process Outcomes and Organizational Agility ..... 509**



Ahmad Syamil (Bina Nusantara University, Indonesia); Fajar Setiadi (PT Telkom Indonesia Tbk, Indonesia); William Valente (STIE, Indonesia)..... 509

**Sustainability of Key Performance Indicators (KPI) Halal Eco-Tourism Information System ..... 514**

Muhammad Syarif Hartawan (University of Krisnadwipayana Indonesia & ICT Departement Asia E University Malaysia, Indonesia); Maya DD Maharani (Sahid University, Indonesia); Erly Krisnanik (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Hoga Saragih (Bakrie University & Indonesia, Indonesia); Aedah Abd Rahman (Universiti Kuala Lumpur, Malaysia)..... 514

**Impact of Service Communication Technology, and Effectiveness on Public Service Performance by Chatbot Service on Ticket Sales Services at PT KAI..... 518**

Tiurida Anita (Bina Nusantara University, Indonesia); Athea Sarastiani (University of Trisakti, Indonesia); Ruddy Awaluddin (Trisakti University, Indonesia) ..... 518

**Digital Technology in Health Education Trend Mapping Through Bibliometrics Analysis..... 523**

Fairuz Iqbal Maulana (Bina Nusantara University, Indonesia); Dian Lestari (Airlangga University, Indonesia); Vandha Widartha (Telkom University, Indonesia); Nirwana Hari (Universitas Madura, Indonesia); Wahyu Dirgantara (University of Merdeka Malang, Indonesia)..... 523

**Energy Saving Method in an Illumination System for Brawijaya Museum Display 529**

Andi Pramono, M. Febriantono, Ida Ananta Wijaya and Tiara Ika Widia Primadani (Bina Nusantara University, Indonesia); Miranti Nurul Huda (Bina Nusantara University & Yayasan Bina Nusantara, Indonesia); Satrio Arif Budiman (University of Brawijaya, Indonesia)..... 529

**The Smart Lighting System in the Coworking Space's Meeting Room..... 534**

Anastasya Krisna Putri (School of Design & BINUS University, Indonesia); Andi Pramono (Bina Nusantara University, Indonesia); Diah Maharani Yasmin (Binus University, Indonesia); Bela Ayu Safitri (BINUS University, Indonesia); Yukiko Safira Zaharani and Felix Franklin Zebua (Binus University, Indonesia)..... 534

**You-Topia: Virtual World Game Mobile App to Improve Creativity ..... 539**

Ivana Clarissa Soegiarto, Yudhistya Ayu Kusumawati and Cuk Tho (Bina Nusantara University, Indonesia) ..... 539

**Factors Influencing Indonesian Consumers to Use Digital Banking ..... 546**

Christine Tanika and William William (BINUS University, Indonesia); Michael Putra Hartanto (Bina Nusantara University, Indonesia); Miranda Tanjung (BINUS University, Indonesia)..... 546



**Systematic Literature Review of Augmented Reality and Virtual Reality Development in Education..... 552**

Jevon Danaristo, Julian Jearsen, Yonathan Faitho, Jurike Moniaga and Bakti Amirul Jabar (Bina Nusantara University, Indonesia)..... 552

**Augmented Reality Assisted Healthy Drinking Water Products Promotion and Literacy: A Development and Evaluation ..... 557**

Athika Dwi Wiji Utami, Trias W Andari, Putra Uji Deva Satrio and Sonhaji Arif (Universitas Nahdlatul Ulama Sidoarjo, Indonesia)..... 557

**A Systematic Review of M-Learning and Entrepreneurship ..... 563**

Agung Purnomo (Bina Nusantara University, Indonesia); Mega Firdaus and Elsa Rosyidah (Universitas Nahdlatul Ulama Sidoarjo, Indonesia); M Meiryani and Fairuz Iqbal Maulana (Bina Nusantara University, Indonesia); Evaristus Didik Madyatmadja (Bina Nusantara University & School of Information Systems, Indonesia) ..... 563

**Improving Education Quality in Cloud Education Environment..... 568**

Kosasih Ali Abu Bakar (Hefei University of Technology & Ministry of Education, Culture, Research, and Technology, Indonesia); Erly Krisnanik and Bambang Saras Yulistiawan (Universitas Pembangunan Nasional Veteran Jakarta, Indonesia); Ren Minglun (Hefei University of Technology, China)..... 568

# Concatenate Word Embedding for Text to Image through Generative Adversarial Network

1<sup>st</sup> Rakhmi Khalida

Department of Informatics Engineering  
Gunadarma University  
Depok, Indonesia  
rakhmikhhalida7@gmail.com

2<sup>nd</sup> Sarifuddin Madenda

Department of Informatics Engineering  
Gunadarma University  
Depok, Indonesia  
sarifuddinmadenda@gmail.com

3<sup>rd</sup> Suryadi Harmanto

Department of Informatics Engineering  
Gunadarma University  
Depok, Indonesia  
suryadi.hs@gunadarma.ac.id

4<sup>th</sup> I Made Wiryana

Department of Informatics Engineering  
Gunadarma University  
Depok, Indonesia  
mwiryana.mobile@gmail.com

**Abstract**— Since the explosion of deep learning, automatic image generation from natural language is highly desirable through artificial intelligence (AI) as it makes it easier for users to create visually rich images through the ease of language. One of the methods used to generate images from text is GAN, in this study using word embedding is to process natural language and develop GAN by concatenate word embedding. In this work, we carry out our early-stage research which is to explore the simple technique of concatenate word embedding as the input of the GAN neural network. We show that this model is a novelty for the GAN model with the concept of multimodal input that is able to generate text to image and is expected to improve performance on a stronger GAN. Based on the explanation of our research method, this model can be implemented and can be developed for various GAN tasks such as style transfer, image to image, face inpainting or image repair semantically, and super resolution.

**Keywords**— text, image, word embedding, GAN

## I. INTRODUCTION

Since the explosion of deep learning has begun, many researchers have begun to build many neural network architectures. It is often speculated that neural networks are inspired by neurons and their networks in the brain. Artificial Neural Network (ANN) algorithms often mimic and copy these biological structures, but much remains to be discovered about how the brain works [1]. The Artificial Neural Network (ANN) Algorithm simply cannot solve the mystery of the brain. That is why artificial intelligence scientists have to invent multiple neural network architectures to accomplish different tasks.

A picture is worth a thousand words, while the text on a picture makes a description or storyline clear. Such abilities are essential in generating creative thinking for the visual arts. Machines cannot have such capabilities, this can be formulated into the text to image. Text to image is the creation of an image from a text description. Text to image is a complex task of computer vision and deep learning which has made great progress over the last few years. Automatic image generation from natural language or generating images from text allows users to visually depict rich elements through text descriptions. Such capabilities are effectively highly desirable to be created through artificial intelligence (AI). Advances in deep learning have resulted in network architectures that promise to produce realistic images.

One of the algorithms that can generate images from text is the Generative Adversarial Network (GAN). GAN is an

artificial neural network created [2], which can generate realistic images or artistic images such as paintings. In recent years, GANs have been developed to perform increasingly complex tasks, such as style transfer[3], [4], image-to-image [5], [6], human faces [7], face inpainting or semantic image repair [8], [9], text to image [10], super-resolution [11], representation learning [12], and data augmentation [13]. The proprietary Deep Convolution Generative Adversarial Network (DC-GAN) architecture [10] pioneered the algorithm for text-to-image using GAN. DC-GAN is a multimodal learning model that tries to bridge the two unsupervised machine learning algorithms, namely RNN and GAN neural networks to generate text to image. The RNN network serves to represent text into vectors for GAN network input. DC- GAN's contribution is to produce reasonable but effective images of natural language. The DC-GAN algorithm was tested on the Caltech-UCSD Birds datasets. Each image in this data set has five text descriptions.

The main work in generating images from text is to represent natural language in the form of text information in vector form and enter it into the GAN neural network. Text representation can generally be expressed in a distributed form, where the semantics are spread into the vector dimensions through matrix dimension reduction or neural network dimension reduction, which is called word embedding.

Mikolov et al from Google [14] released the word vector training toolkit Word2vec for expressing word semantic information. Word2vec is a static word embedding model that can get words from a predicted local context by maximizing conditional probabilities. Pennington et al. [14] launched a word vector model with the name Global Vector (GloVe) to create word vectors based on global information. Glove and word2vec is a static word embedding that can be used to handle word similarity, word analogy, and named entity recognition.

Previous studies often used single static word vectors which could not take into account multipart feature information from sentences. In our research, we propose a way to build a more robust GAN that can generate images from text by using the concatenate word embedding method. This research is still in its early stages, the result of this research is a word embedding design that is combined as input from the GAN neural network. We show that this model is a novelty for the GAN model with the concept of multimodal input capable of generating text to image. This research use dataset is grouped into 8K images for training and 3K images for testing.

---

This work was supported in full by Indonesia Endowment Fund for Education (LPDP) Ministry of Finance Republic Indonesia.

This paper is structured as follows, section 2 contains similar work done with word embedding and GAN networks that can generate images from text. We present word2vec and glove approach methods and the GAN architecture in section 3, In Section 4, we describe the experimental setup and the data set used. The results of this study are presented in Section 5, and section 6 is the conclusions of this paper.

## II. RELATED WORK

The development of GPU computing and significant advances in hardware performance has pushed the task of generating images from text with GAN deep learning technology to attract people's attention. Since GAN's inception in 2014 [1], the task of generating images from text with GANs has attracted more and more attention, and many GAN-based models have been developed.

In general, the GAN architecture consists of 2 networks which are referred to as generator and discriminator networks. The shape of the generator network can be seen in contrast to the structure of the neural network in general. The generator network uses a deconvolutional layer whose task is to change the dimensions of the input data into a larger dimension (Upsampling). The output of this generator is a 3-dimensional matrix which will then be referred to as data generation. A generator network usually consists of a fully connected layer, dense layer, batch normalization, Leaky ReLU, and reshape layer.

The discriminator network is a binary classification network that accepts three-dimensional image input and outputs a classification value of 0 or 1. The classification of the discriminator is an assessment of the generator in generating the generated image. If the generated image is like the original dataset then the discriminator value is 1 and vice versa.

The way GAN works to create generative images is through the adversarial process. Two models, namely generator G and discriminator D are designed to compete with each other. G tries to imitate generating a sample distribution, while D tries to judge the data comes from G or which is sample data of the dataset. The results of the assessment are then calculated with an optimizer function or loss function to optimize the network. Both can learn from the results. This process is executed many times. According to research [1], it is possible for the probability value of D is 0.5 to assess the data from the sample or from the results generator.



Fig. 1. The Step of Original GAN

Reed et al, were the first to develop GAN with the aim of generating images from text. His research uses RNN network to convert text input into character level so that it becomes code. Another contribution of his research is pairing original images with mismatched text other than original images with matching text and images generate with matching text. It aims to be able to generate images outside of the context of text input.

Another architecture developed by Gorti et al created MirrorGAN. The idea is to extend the cycleGAN architecture and

add stacking, attention, and cycle consistency. MirrorGAN also uses word embedding to convert text input into word vectors, that code that feeds into the GAN generator [15].

Zhang et al built StackGAN by stacking two GANs to generate images from text. In his research, the stage-I GAN produced a low-resolution outline of the resulting image from the desired text and the stage-II GAN filled in the details of the image sketch. Another contribution from Zhang et al. is that the fully connected layer produces the mean and variance before taking text input and samples from the normal distribution. This sampling can add data and increase robustness [16].

The use of GAN and word embedding to generate text has been explored by Rajeswar et al [17]. This work tells the story by using embeddings at the sentence level allowing the model to represent the expression of an author in creating a text in a simple and effective way, rather than using word-level embedding.

Two important and widely used word embedding models because they have significant results in downstream NLP tasks are word2vec and glove embedding. Naderalvojud et al. [18] applied pre-trained GloVe and Word2vec embedding vectors in their deep learning model and improved sentiment classification accuracy. Rezaeimia et al. [19] proposed to combine several word vector representations such as Word2vec, GloVe, POS2Vec, Lexicon2Vec, and Word-position2Vec to verify the accuracy of sentiment prediction.

Text-to-Face (TTF) synthesis has been carried out. In this paper, we propose a Text-to-Face model that not only produces images in high resolution (1024x1024) with text-to-image consistency and also displays multiple multifaceted faces to cover various facial features that are not naturally defined [20].

## III. METHOD

The three stages carried out in this research are the stages of data loading, training, and generates. The steps of proposed system GAN can be seen in Fig 2.

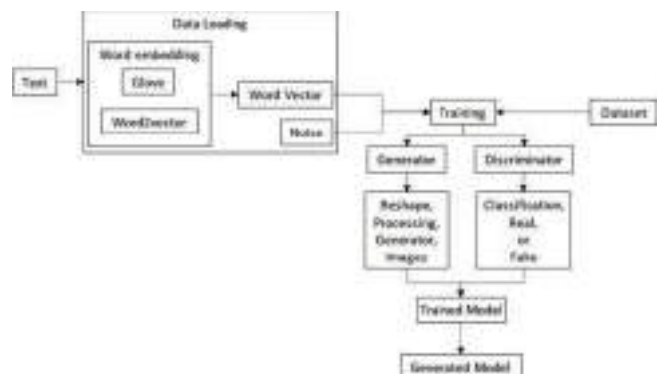


Fig. 2. Steps of Proposed System

In our formulation, not only noise as input to Generator, textual description is first converted into word vector by text insertion, then combined with random noise vector as input to Generator. For example, the textual description has been converted to a 256-dimensional insert and merged with a 100-dimensional noise vector [sampled from the Normal distribution. This formulation will help the Generator to generate an image that matches the input description and does not generate random images.

A. Data Loading

At the data loading stage, the text description is converted into word vector by word embedding, and the image data is converted into an array using a trained network such as Inception V3 CNN. Changes in text data into word vectors through the word embedding model, where each word is mapped to a different vector form. Word vector can be seen in fig 3.



Fig. 3. Representation of Word Vector

B. Training

The training stage is the stage of the system recognizing image and text objects. Generators and discriminators are trained together, each of which may consist of a network layer consisting of a ResNet network (Residual Network). The training process for a GAN means training the generator and discriminator simultaneously. The generator work to generate images that match the text description provided, while the discriminator evaluates the generated images based on the input of three pairs, namely the original image with matching text, generated image with matching text, and the original image with unmatched text. The process of training GAN is not easy, in training that uses adversarial and backpropagation methods, neither generator nor discriminator can win. When the generator wins from the discriminator there will be overfitting or producing the same generated image. When the discriminator wins from the generator, there will be no learning process from the generator. Hyperparameters on the GAN architecture are very important to set in the training process. In the training process using learning rate, batch, and epoch.

C. Generate

The generate stage is the stage after getting a trained model and is able to produce images according to text input, whether text is in the dataset or not. At this stage the system will create a checkpoint that is stored in the file directory. This checkpoint file can be reused someday to continue train GAN network with parameter update or not. The system will also save the Generator and Discriminator models in a predefined format such as h5. When this step is over, you can evaluate whether the GAN is performing well enough.

IV. DATA, TOOLS, AND ENVIRONMENT

This study using hardware with specifications for an Intel Core i7 processor, clock speed of 2.3 GHz, Nvidia Geforce RTX 3060 GPU engine graphics, 16GB RAM, and Windows 11 operating system with 64bit architecture. The software specifications used include python version 3, anaconda, jupyter, and google colab for the editor, then Pytorch, TensorFlow, tensorlayer, hard, and the python library.

The image and text description dataset used is Caltech-UCSD, the dataset overall contains 11k images of birds that have been collected from the labeled bird in the wild. The dataset is grouped into 8K images for training and 3K images

for testing, which have been described by special annotators, and the corpus dataset is used as word embedding to change the text description. to be word vector is Wikipedia and Gigaword 5 vector with 300 dimensions.

V. IMPLEMENTATION

A. Word Embedding

Word embedding is one of the Natural Language Processing (NLP) techniques in which words or phrases i.e., strings of vocabulary are mapped onto a vector of real numbers. The need to map strings to vectors of real numbers stems from the computer’s inability to perform operations with strings. Since the one hot encoding method is inefficient and unique ID does not offer a relational representation, word embedding is able to see the relational representation of words and can represent each word with a vector that is simpler than one hot encoding. The final array size is also much smaller than the one hot encoding encoded vocabulary.

Word embedding can adopt a supervised, unsupervised or semi-supervised learning approach. The word embedding model that will be used is Word2vec and glove. The Word2vec and glove models will create an embedding matrix for all words in the corpus. The word2vec model contains a matrix of vector length with a fixed vector size while glove builds a large matrix by taking into account the global word for word count in the corpus.

Word2vec and glove are word embedding types of static words with an unsupervised approach. This is because to find the ideal vector value for each word and usually use language models to predict the next word based on its context. The word embedding structure can be seen in the Fig 4.

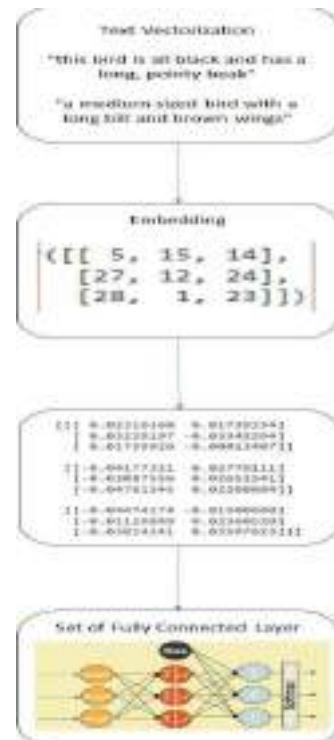


Fig. 4. Word Embedding Structure

The explanation of the word embedding structure in the image above is as follows:



- 1) *Text Vectorization Layer* to convert text to 1D integer tensor This layer performs the following procedure:
  - a) *Change every uppercase letter* in a sentence to lowercase.
  - b) *Remove every punctuation.*
  - c) *Turning sentences into words.*
  - d) *Merge words into tokens.*
  - e) *Doing token index.*
  - f) *Converts each sentence using an index into a 1D tensor of an integer token index or float value.*

2) *Embedding layer* to convert 1D integer tensor to solid vector with a fixed size.

3) *Neural network* fully connected layers with backpropagation algorithm and loss function as well as other networks. The values that have been obtained are still very random. Therefore, we need an optimizer to calculate the loss function and adjust its value with backpropagation and this optimizer must run on a set of layers called fully connected layers such as LSTM, GRU, or Convolution layers. Using a set of fully connected layers can insert strings into labels to match vector values.

**B. Concatenate Word Embedding**

First, in the word2vec model, training was conducted on the Wikipedia corpus with skip-gram architecture, 300 dimensions in size, 3 threads for multiprocessing, and window size 3. The use of a smaller window will give a more specific term to a word context. The frequency of occurrence of words or called min count is 1. Words that rarely appear are usually not very important, so the model will ignore words that do not meet the min-count value.

The word embedding glove model was trained on a Gigaword 5 vector corpus with memory parameters is 4, verbose 2, size 300, and binary is 2. Verbose is a value that tells the function how much information to print when training the model, the larger the value, the more detail. generated information. Memory in glove and thread in word2vec both serve as a measure of memory usage used during model training. Binary is an option to output the model, a value of 0 means text output type, 1 value for binary output type, and 2 value for text and binary output. This parameter is more influential for the efficiency of memory usage and model size. Both word embedding models are trained using CNN networks because using a set of fully connected layers can insert strings into labels to match vector values. and the backpropagation algorithm is needed as an optimizer that calculates the loss function and adjust its value to produce a better semantic value.

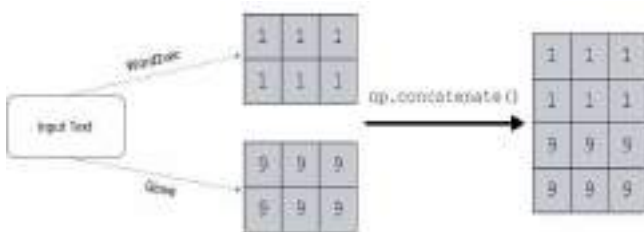


Fig. 5. Illustration of Concat Word Embedding

Inspired by the form of a word vector is the fixed array, for example, 50, 100, or 300 dimensions, which when con-

catenate with the model, number of arrays becomes the total dimensions of the sum of the two models. The process of concatenate two models word embedding into one model can be seen in the Fig. 5.

**C. Text to Image with GAN**

The GAN architecture for this model consists of a GAN neural network that has a multimodal input consisting of images and text. The combination of word embedding as a text modifier and together with images becomes input for the generator and will be judged genuine or fake by the discriminator in the GAN.

Details of the GAN generator network consists is 5 layers of deconvolution layer. The generator works to change the dimensions of the input data into a larger dimension. The generator consisting of a 2-dimensional deconvolution layer receives input in is noise, this noise will go through a dense layer, batch normalization, RELU, and the output layer is a reshape layer. This is repeated in up to 4 parts. The output of this generator is a 3-dimensional matrix which will then be referred to as data generation. An image of the structure of the generator can be seen in Fig 6.



Fig. 6. Structure of Generator

The discriminator consists of 5 convolutional layers, the image that is input to the discriminator will be entered in the first part consisting of a 2-dimensional convolutional layer. Leaky RELU, and drop out, then at the end, the data will go through a flattened, dense, fully connected layer, so that it produces 1 value in the form of a decision. If the data entered is an image from the dataset, the discriminator produces a value close to 1, vice versa. Image of the structure of the discriminator can be seen as fig 7.



Fig. 7. Structure of Discriminator

Detailed image of GAN architecture with multimodal input using concatenate word embedding is described as fig 8.

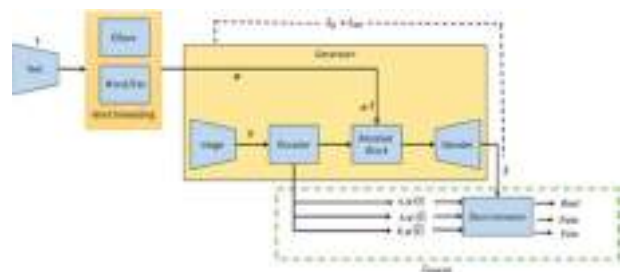


Fig. 8. Proposed Structure of GAN

In the process of training the GAN network, there is a loss function that functions to measure the balance between the power of the generator and the power of the discriminator. it's a min-max optimization formulation where Generator wants



to minimize the objective function whereas Discriminator wants to maximize the same objective function. The weakness in the GAN that often occurs is the collapse or overfitting mode. The loss function generator is made maximal so that it does not lose with the discriminator loss function.

This min-max formulation of the objective function has a global optimum. The loss function in this research adopts AttnGAN's property[21]. DAMSM loss function is able to calculate the similarity between the resulting image and the sentence using global sentence level information (t) and detailed word level information (w). DAMSM loss function measures the degree of match between the image and the text description. Based on the DAMSM loss function, the discriminator classification the image according to 3 categories:

- 1) Original image with matching text  $D(x, \varphi(t))$
- 2) Generated image with matching text  $D(x^\wedge, \varphi(\bar{t}))$
- 3) Original image with mismatched text  $D(x, \varphi(\bar{t}))$

The described GAN architecture has the basic parameters of DCGAN [10] and adopts AttnGAN's DAMSM loss function [10]. DAMSM loss function is designed to study model attention in a semi-supervised manner, where the only supervision is matching between the all of picture and the all of sentence (order word). A collection of pairs of pictures-description  $(Q_i, D_i)^M$ . Posterior the probability of a  $D_i$  sentence being paired with a  $Q_i$  image is calculated as

$$P(Q_i, D_i) = \frac{\exp(\gamma R(Q_i, D_i))}{\sum_{i=1}^M \exp(\gamma R(Q_i, D_i))} \quad (1)$$

Substitute value for  $R(Q, D)$  is

$$R(Q_i, D_i) = \log \sum_{i=1}^{T-1} \exp(\gamma R(c_i, e_i))^{\gamma^2} \quad (2)$$

The following is loss function formula for DAMSM based (1) and (2)

$$L_{D,E,MSM} = L^W + L^W + L^t + L^t \quad (3)$$

when,

$$L^w = - \sum_{i=1}^M \log P(Q_i | D_i) \quad (4)$$

Also applies to  $L^w$  and if (2) defined with  $R = (Q|D) = \frac{(\sqrt{T}e)}{(|\bar{v}| || \bar{e}|)}$  then substitution to (1) and (4) is obtained  $L_1^t$  and  $L_2^t$ .

The following is loss function formula for generator:

$$L_G = - \frac{1}{2} [E_{x-pG} \log(D(x)) + E_{x-pG} \log(D(x, \varphi(t)))] \quad (5)$$

When the input is a noise distribution, many probabilities are generated due to vector differences. Some exist mode is preferred over other modes. The mode-seeking loss function tries to amplify the distribution of minor modes between major modes. When the noise distribution is used to generate two different samples based on the same input. The first point  $z_1$  causes the formation of a mode image  $I_1 = G(c, z_1)$  and the second point  $z_2$  points to the image  $I_2 = G(c, z_2)$  after applying generate image and text (c).  $D(\cdot, \cdot)$  is a distance function that calculates the distance between two tensors or vectors. The numerator calculates the distance between object vectors while the denominator is the distance between ordinary vectors. The purpose of this formula is to maximize the distance between vector objects. The following is the MSGAN loss function formula.

$$L_{MS} = \max G \left( \frac{DI(G(c, Z_1), G(c, Z_2))}{D_Z(Z_1, Z_2)} \right) \quad (6)$$

Based on (1) until (6), The calculation of the loss function formula is (7)

$$L = \sum_i L_G + L_{D,E,MSM} + l_{MS} \quad (7)$$

The structure of the GAN architecture that adopts DCGAN [10], the difference between the input generators is the distribution of noise and text. The text encoder does not use word embedding but a Skip-thought model[22]. Adoption of the DCGAN structure as it mimics the workings of discriminator networks studying the appropriate relationship between description of text and image content with GAN-CLS and GAN-INT function.

At the start of the training, the discriminator ignores the text and easily reject the image generated from the generator, because it doesn't look reasonable, but after the generator receives feedback update weight from the discriminator, the generator learns to generate a reasonable generated image and learns to align it with the text, the the discriminator also learn to evaluate the resulting image to satisfy this conditioning problem. Based on this intuition it may complicate the dynamics of learning, but GAN-CLS pays attention to this by adding an alternate decision to the discriminator, namely a third input consisting of the original image with mismatched text. It is in accordance with the DAMSM loss function.

GAN-INT works with word embedding on the generator side. GAN-INT utilizes manifold interpolation learning to generate additional text for word embedding during training. This matter serves to make the generator stronger to generate images and can deceive the discriminators from being able to distinguish the output. The results of the added text to be inserted do not need to match the facts so that GAN-INT does not require a cost.

## VI. CONCLUSION

Word embedding glove and word2vec are static word models which are the most widely used word embedding. The availability of pretrained word embedding models and with vocabulary area coverage can reduce the high computational requirements for training word embedding models. In this work, we explore a simple technique, namely concatenate word embedding, which contributes to developing GAN for the task of generating images from text and is expected to improve performance on a more robust GAN. Based on the explanation of our research method, this model can be implemented and can be developed for various GAN tasks such as style transfer, image to image, face inpainting or image repair semantically, and super resolution.

## ACKNOWLEDGMENT

Thank to Indonesia Endowment Fund for Education (LPDP) Ministry of Finance Republic Indonesia have been made possible this research and presentation.

## REFERENCES

- [1] Martin C Nwadiugwu, "Neural Networks, Artificial Intelligence and the Computational Brain," *ArXiv*, 2021.
- [2] Ian J. Goodfellow *et al.*, "Generative Adversarial Nets," 2014.
- [3] Y. Jing, Y. Yang, Z. Feng, J. Ye, and M. Song, "Neural style transfer," *IEEE Trans Vis Comput Graph*, 2019.
- [4] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2016, pp. 2414–2423.

- [5] J.-Y. Z. T. Z. A. A. E. P. Isola, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2016, pp. 1125–1134.
- [6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.
- [7] T. A. S. L. J. L. T. Karras, "Progressive growing of gans for improved quality, stability, and variation," 2018.
- [8] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Free- form image inpainting with gated convolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp.4471–4480.
- [9] R. A. Yeh, C. Chen, T.-Y. Lim, A. G. Schwing, and M. N. M. Hasegawa Johnson, "Semantic image inpainting with deep generative models," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2016, pp. 5485–5493.
- [10] Scott Reed and Zeynep Akata and Xinchun Yan and Lajanugen Logeswaran and Bernt Schiele and Honglak Lee, "Generative Adversarial Text to Image Synthesis," *Proc Mach Learn Res*, 2016.
- [11] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2016, pp. 4681–4690.
- [12] J. Donahue and K. Simonyan, "Large scale adversarial representation learning," *Advances in Neural Information Processing Systems*, pp. 10452–10552, 2019.
- [13] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," 2013.
- [14] R. S. and C. D. M. J. Pennington, "GloVe : Global Vectors for Word Representation," 2014.
- [15] S. K. Gorti and J. Ma, "Text-to-Image-to-Text Translation using Cycle Consistent Adversarial Networks," Aug. 2018, [Online]. Available: <http://arxiv.org/abs/1808.04538>
- [16] Han Zhang and Tao Xu and Hongsheng Li and Shaoting Zhang and Xiaogang Wang and Xiaolei Huang and Dimitris Metaxas, "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks." International Conference on Computer Vision (ICCV), 2017.
- [17] Sai Rajeswar, Sandeep Subramanian, Francis Dutil, Christopher Joseph Pal, and Aaron C. Courville, "Adversarial generation of natural language," *CoRR*, 2017.
- [18] Behzad Naderalvojud and Ebru Akcapinar Sezer, "Sentiment aware word embeddings using refinement and senti-contextualized learning approach," *Neurocomputing*, vol. 405, pp. 149–160, 2020.
- [19] Seyed Mahdi Rezaeinaia, Rouhollah Rahmania, Ali Ghodsib, and Hadi Veisia, "Sentiment analysis based on improved pre-trained word embeddings," *Expert Syst Appl*, vol. 117, pp. 139–147, 2019.
- [20] A. D. Pavlo, Lucchi, and T. Hofmann, "Faces à la Carte: Text-to-Face Generation via Attribute Disentanglement," *European Conference on Computer Vision*, pp. 482–499, 2020.
- [21] 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," *CoRR*, 2017.
- [22] Ryan Kiros *et al.*, "Skip-Thought Vectors," *CoRR*, 2015.

# Concatenate Word Embedding for Text to Image through Generative Adversarial Network

1<sup>st</sup> Rakhmi Khalida  
Department of Informatics Engineering  
Gunadarma University  
Depok, Indonesia  
rakhmikhalida7@gmail.com

2<sup>nd</sup> Sarifuddin Madenda  
Department of Informatics Engineering  
Gunadarma University  
Depok, Indonesia  
sarifuddinmadenda@gmail.com

3<sup>rd</sup> Suryadi Hamanto  
Department of Informatics Engineering  
Gunadarma University  
Depok, Indonesia  
suryadi.hs@gunadarma.ac.id

4<sup>th</sup> I Made Wiryana  
Department of Informatics Engineering  
Gunadarma University  
Depok, Indonesia  
imwiryana.mobile@gmail.com

**Abstract**— Since the explosion of deep learning, automatic image generation from natural language is highly desirable through artificial intelligence (AI) as it makes it easier for users to create visually rich images through the ease of language. One of the methods used to generate images from text is GAN, in this study using word embedding is to process natural language and develop GAN by concatenate word embedding. In this work, we carry out our early-stage research which is to explore the simple technique of concatenate word embedding as the input of the GAN neural network. We show that this model is a novelty for the GAN model with the concept of multimodal input that is able to generate text to image and is expected to improve performance on a stronger GAN. Based on the explanation of our research method, this model can be implemented and can be developed for various GAN tasks such as style transfer, image to image, face inpainting or image repair semantically, and super resolution.

**Keywords**— text, image, word embedding, GAN

## I. INTRODUCTION

Since the explosion of deep learning has begun, many researchers have begun to build many neural network architectures. It is often speculated that neural networks are inspired by neurons and their networks in the brain. Artificial Neural Network (ANN) algorithms often mimic and copy these biological structures, but much remains to be discovered about how the brain works [1]. The Artificial Neural Network (ANN) Algorithm simply cannot solve the mystery of the brain. That is why artificial intelligence scientists have to invent multiple neural network architectures to accomplish different tasks.

A picture is worth a thousand words, while the text on a picture makes a description or storyline clear. Such abilities are essential in generating creative thinking for the visual arts. Machines cannot have such capabilities, this can be formulated into the text to image. Text to image is the creation of an image from a text description. Text to image is a complex task of computer vision and deep learning which has made great progress over the last few years. Automatic image generation from natural language or generating images from text allows users to visually depict rich elements through text descriptions. Such capabilities are effectively highly desirable to be created through artificial intelligence (AI). Advances in deep learning have resulted in network architectures that promise to produce realistic images.

One of the algorithms that can generate images from text is the Generative Adversarial Network (GAN). GAN is an

artificial neural network created [2], which can generate realistic images or artistic images such as paintings. In recent years, GANs have been developed to perform increasingly complex tasks, such as style transfer[3], [4], image-to-image [5], [6], human faces [7], face inpainting or semantic image repair [8], [9], text to image [10], super-resolution [11], representation learning [12], and data augmentation [13]. The proprietary Deep Convolution Generative Adversarial Network (DC-GAN) architecture [10] pioneered the algorithm for text-to-image using GAN. DC-GAN is a multimodal learning model that tries to bridge the two unsupervised machine learning algorithms, namely RNN and GAN neural networks to generate text to image. The RNN network serves to represent text into vectors for GAN network input. DC-GAN's contribution is to produce reasonable but effective images of natural language. The DC-GAN algorithm was tested on the Caltech-UCSD Birds datasets. Each image in this data set has five text descriptions.

The main work in generating images from text is to represent natural language in the form of text information in vector form and enter it into the GAN neural network. Text representation can generally be expressed in a distributed form, where the semantics are spread into the vector dimensions through matrix dimension reduction or neural network dimension reduction, which is called word embedding.

Mikolov et al from Google [14] released the word vector training toolkit Word2vec for expressing word semantic information. Word2vec is a static word embedding model that can get words from a predicted local context by maximizing conditional probabilities. Pennington et al. [14] launched a word vector model with the name Global Vector (GloVe) to create word vectors based on global information. Glove and word2vec is a static word embedding that can be used to handle word similarity, word analogy, and named entity recognition.

Previous studies often used single static word vectors which could not take into account multipart feature information from sentences. In our research, we propose a way to build a more robust GAN that can generate images from text by using the concatenate word embedding method. This research is still in its early stages, the result of this research is a word embedding design that is combined as input from the GAN neural network. We show that this model is a novelty for the GAN model with the concept of multimodal input capable of generating text to image. This research use dataset is grouped into 8K images for training and 3K images for testing.

This work was supported in full by Indonesia Endowment Fund for Education (LPDP) Ministry of Finance Republic Indonesia.



This paper is structured as follows, section 2 contains similar work done with word embedding and GAN networks that can generate images from text. We present word2vec and glove approach methods and the GAN architecture in section 3. In Section 4, we describe the experimental setup and the data set used. The results of this study are presented in Section 5, and section 6 is the conclusions of this paper.

## II. RELATED WORK

The development of GPU computing and significant advances in hardware performance has pushed the task of generating images from text with GAN deep learning technology to attract people's attention. Since GAN's inception in 2014 [1], the task of generating images from text with GANs has attracted more and more attention, and many GAN-based models have been developed.

In general, the GAN architecture consists of 2 networks which are referred to as generator and discriminator networks. The shape of the generator network can be seen in contrast to the structure of the neural network in general. The generator network uses a deconvolutional layer whose task is to change the dimensions of the input data into a larger dimension (Upsampling). The output of this generator is a 3-dimensional matrix which will then be referred to as data generation. A generator network usually consists of a fully connected layer, dense layer, batch normalization, Leaky ReLU, and reshape layer.

The discriminator network is a binary classification network that accepts three-dimensional image input and outputs a classification value of 0 or 1. The classification of the discriminator is an assessment of the generator in generating the generated image. If the generated image is like the original dataset then the discriminator value is 1 and vice versa.

The way GAN works to create generative images is through the adversarial process. Two models, namely generator G and discriminator D are designed to compete with each other. G tries to imitate generating a sample distribution, while D tries to judge the data comes from G or which is sample data of the dataset. The results of the assessment are then calculated with an optimizer function or loss function to optimize the network. Both can learn from the results. This process is executed many times. According to research [1], it is possible for the probability value of D is 0.5 to assess the data from the sample or from the results generator.



Fig. 1. The Step of Original GAN

Reed et al, were the first to develop GAN with the aim of generating images from text. His research uses RNN network to convert text input into character level so that it becomes code. Another contribution of his research is pairing original images with mismatched text other than original images with matching text and images generate with matching text. It aims to be able to generate images outside of the context of text input.

Another architecture developed by Gori et al created MirrorGAN. The idea is to extend the cycleGAN architecture and

add stacking, attention, and cycle consistency. MirrorGAN also uses word embedding to convert text input into word vectors, that code that feeds into the GAN generator [15].

Zhang et al built StackGAN by stacking two GANs to generate images from text. In his research, the stage-I GAN produced a low-resolution outline of the resulting image from the desired text and the stage-II GAN filled in the details of the image sketch. Another contribution from Zhang et al. is that the fully connected layer produces the mean and variance before taking text input and samples from the normal distribution. This sampling can add data and increase robustness [16].

The use of GAN and word embedding to generate text has been explored by Rajeswar et al [17]. This work tells the story by using embeddings at the sentence level allowing the model to represent the expression of an author in creating a text in a simple and effective way, rather than using word-level embedding.

Two important and widely used word embedding models because they have significant results in downstream NLP tasks are word2vec and glove embedding. Naderalvojud et al. [18] applied pre-trained GloVe and Word2vec embedding vectors in their deep learning model and improved sentiment classification accuracy. Rezaeinia et al. [19] proposed to combine several word vector representations such as Word2vec, GloVe, POS2Vec, LEXICON2Vec, and Word-position2Vec to verify the accuracy of sentiment prediction.

Text-to-Face (TTF) synthesis has been carried out. In this paper, we propose a Text-to-Face model that not only produces images in high resolution (1024x1024) with text-to-image consistency and also displays multiple multifaceted faces to cover various facial features that are not naturally defined [20].

## III. METHOD

The three stages carried out in this research are the stages of data loading, training, and generates. The steps of proposed system GAN can be seen in Fig 2.

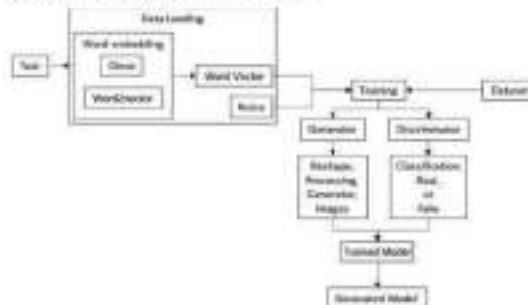


Fig. 2. Steps of Proposed System

In our formulation, not only noise as input to Generator, textual description is first converted into word vector by text insertion, then combined with random noise vector as input to Generator. For example, the textual description has been converted to a 256-dimensional insert and merged with a 100-dimensional noise vector [sampled from the Normal distribution. This formulation will help the Generator to generate an image that matches the input description and does not generate random images.

#### A. Data Loading

At the data loading stage, the text description is converted into word vector by word embedding, and the image data is converted into an array using a trained network such as Inception V3 CNN. Changes in text data into word vectors through the word embedding model, where each word is mapped to a different vector form. Word vector can be seen in fig 3.



Fig. 3. Representations of Word Vector

#### B. Training

The training stage is the stage of the system recognizing image and text objects. Generators and discriminators are trained together, each of which may consist of a network layer consisting of a ResNet network (Residual Network). The training process for a GAN means training the generator and discriminator simultaneously. The generator work to generate images that match the text description provided, while the discriminator evaluates the generated images based on the input of three pairs, namely the original image with matching text, generated image with matching text, and the original image with unmatched text. The process of training GAN is not easy, in training that uses adversarial and backpropagation methods, neither generator nor discriminator can win. When the generator wins from the discriminator there will be overfitting or producing the same generated image. When the discriminator wins from the generator, there will be no learning process from the generator. Hyperparameters on the GAN architecture are very important to set in the training process. In the training process using learning rate, batch, and epoch.

#### C. Generate

The generate stage is the stage after getting a trained model and is able to produce images according to text input, whether text is in the dataset or not. At this stage the system will create a checkpoint that is stored in the file directory. This checkpoint file can be reused someday to continue train GAN network with parameter update or not. The system will also save the Generator and Discriminator models in a predefined format such as h5. When this step is over, you can evaluate whether the GAN is performing well enough.

### IV. DATA, TOOLS, AND ENVIRONMENT

This study using hardware with specifications for an Intel Core i7 processor, clock speed of 2.3 GHz, Nvidia Geforce RTX3060 GPU engine graphics, 16GB RAM, and Windows 11 operating system with 64bit architecture. The software specifications used include python version 3, anaconda, jupyter, and google colab for the editor, then Pytorch, TensorFlow, tensorlayer, hard, and the python library.

The image and text description dataset used is Caltech-UCSD, the dataset overall contains 11k images of birds that have been collected from the labeled bird in the wild. The dataset is grouped into 8K images for training and 3K images

for testing, which have been described by special annotators, and the corpus dataset is used as word embedding to change the text description to be word vector is Wikipedia and Gigaword 5 vector with 300 dimensions

### V. IMPLEMENTATION

#### A. Word Embedding

Word embedding is one of the Natural Language Processing (NLP) techniques in which words or phrases i.e., strings of vocabulary are mapped onto a vector of real numbers. The need to map strings to vectors of real numbers stems from the computer's inability to perform operations with strings. Since the one-hot encoding method is inefficient and unique ID does not offer a relational representation, word embedding is able to see the relational representation of words and can represent each word with a vector that is simpler than one-hot encoding. The final array size is also much smaller than the one-hot encoding encoded vocabulary.

Word embedding can adopt a supervised, unsupervised or semi-supervised learning approach. The word embedding model that will be used is Word2vec and glove. The Word2vec and glove models will create an embedding matrix for all words in the corpus. The word2vec model contains a matrix of vector length with a fixed vector size while glove builds a large matrix by taking into account the global word for word count in the corpus.

Word2vec and glove are word embedding types of static words with an unsupervised approach. This is because to find the ideal vector value for each word and usually use language models to predict the next word based on its context. The word embedding structure can be seen in the Fig 4.

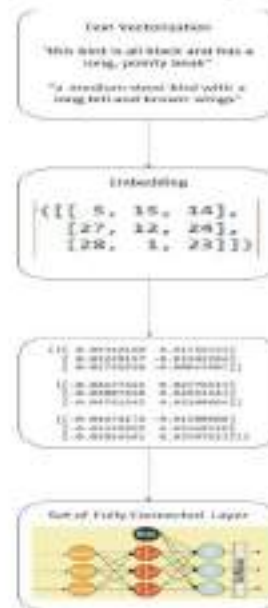


Fig. 4. Word Embedding Structure

The explanation of the word embedding structure in the image above is as follows:



1) *Text Vectorization Layer* to convert text to 1D integer tensor. This layer performs the following procedure:

a) *Change every uppercase letter* in a sentence to lowercase.

b) *Remove every punctuation*.

c) *Turning sentences into words*.

d) *Merge words into tokens*.

e) *Doing token index*.

f) *Converts each sentence using an index into a 1D tensor of an integer token index or float value*.

2) *Embedding layer* to convert 1D integer tensor to solid vector with a fixed size.

3) *Neural network* fully connected layers with backpropagation algorithm and loss function as well as other networks. The values that have been obtained are still very random. Therefore, we need an optimizer to calculate the loss function and adjust its value with backpropagation and this optimizer must run on a set of layers called fully connected layers such as LSTM, GRU, or Convolution layers. Using a set of fully connected layers can insert strings into labels to match vector values.

### B. Concatenate Word Embedding

First, in the word2vec model, training was conducted on the Wikipedia corpus with skip-gram architecture, 300 dimensions in size, 3 threads for multiprocessing, and window size 3. The use of a smaller window will give a more specific term to a word context. The frequency of occurrence of words or called min count is 1. Words that rarely appear are usually not very important, so the model will ignore words that do not meet the min-count value.

The word embedding glove model was trained on a Giga-word 5 vector corpus with memory parameters is 4, verbose 2, size 300, and binary is 2. Verbose is a value that tells the function how much information to print when training the model, the larger the value, the more detail, generated information. Memory in glove and thread in word2vec both serve as a measure of memory usage used during model training. Binary is an option to output the model, a value of 0 means test output type, 1 value for binary output type, and 2 value for test and binary output. This parameter is more influential for the efficiency of memory usage and model size. Both word embedding models are trained using CNN networks because using a set of fully connected layers can insert strings into labels to match vector values, and the backpropagation algorithm is needed as an optimizer that calculates the loss function and adjust its value to produce a better semantic value.

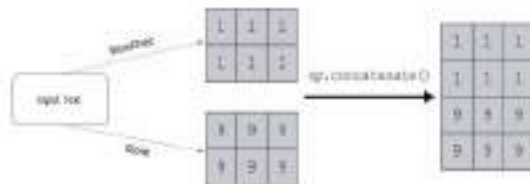


Fig. 5. Illustration of Concat Word Embedding

Inspired by the form of a word vector is the fixed array, for example, 50, 100, or 300 dimensions, which when con-

catenate with the model, number of arrays becomes the total dimensions of the sum of the two models. The process of concatenate two models word embedding into one model can be seen in the Fig. 5.

### C. Text to Image with GAN

The GAN architecture for this model consists of a GAN neural network that has a multimodal input consisting of images and text. The combination of word embedding as a text modifier and together with images becomes input for the generator and will be judged genuine or fake by the discriminator in the GAN.

Details of the GAN generator network consists is 5 layers of deconvolution layer. The generator works to change the dimensions of the input data into a larger dimension. The generator consisting of a 2-dimensional deconvolution layer receives input in is noise, this noise will go through a dense layer, batch normalization, RELU, and the output layer is a reshape layer. This is repeated in up to 4 parts. The output of this generator is a 3-dimensional matrix which will then be referred to as data generation. An image of the structure of the generator can be seen in Fig 6.



Fig. 6. Structure of Generator

The discriminator consists of 5 convolutional layers, the image that is input to the discriminator will be entered in the first part consisting of a 2-dimensional convolutional layer, Leaky RELU, and drop out, then at the end, the data will go through a flattened, dense, fully connected layer, so that it produces 1 value in the form of a decision. If the data entered is an image from the dataset, the discriminator produces a value close to 1, vice versa, image of the structure of the discriminator can be seen as fig 7.



Fig. 7. Structure of Discriminator

Detailed image of GAN architecture with multimodal input using concatenate word embedding is described as fig 8.

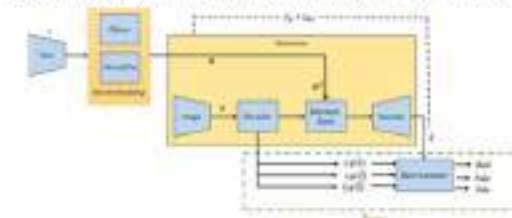


Fig. 8. Proposed Structure of GAN

In the process of training the GAN network, there is a loss function that functions to measure the balance between the power of the generator and the power of the discriminator, it's a min-max optimization formulation where Generator wants

to minimize the objective function whereas Discriminator wants to maximize the same objective function. The weakness in the GAN that often occurs is the collapse or overfitting mode. The loss function generator is made maximal so that it does not lose with the discriminator loss function.

This min-max formulation of the objective function has a global optimum. The loss function in this research adopts AttnGAN's property[21]. DAMSM loss function is able to calculate the similarity between the resulting image and the sentence using global sentence level information (g) and detailed word level information (w). DAMSM loss function measures the degree of match between the image and the text description. Based on the DAMSM loss function, the discriminator classification the image according to 3 categories:

- 1) Original image with matching text  $D(x, \varphi(t))$
- 2) Generated image with matching text  $D(x', \varphi(\bar{t}))$
- 3) Original image with mismatched text  $D(x, \varphi(\bar{t}))$

The described GAN architecture has the basic parameters of DCGAN [10] and adopts AttnGAN's DAMSM loss function [10]. DAMSM loss function is designed to study model attention in a semi-supervised manner, where the only supervision is matching between the all of picture and the all of sentence (order word). A collection of pairs of picture-description  $(Q_i, D_i)^N$ . Posterior the probability of a  $D_i$  sentence being paired with a  $Q_i$  image is calculated as

$$P(Q_i, D_i) = \frac{\exp(\gamma R(Q_i, D_i))}{\sum_{j=1}^N \exp(\gamma R(Q_j, D_i))} \quad (1)$$

Substitute value for  $R(Q, D)$  is

$$R(Q, D) = \log \sum_{i=1}^n \exp(\gamma R(c_i, e_i))^\alpha \quad (2)$$

The following is loss function formula for DAMSM based (1) and (2)

$$L_{DAMSM} = L^w + L^g + L' + L'' \quad (3)$$

when,

$$L' = - \sum_{i=1}^N \log P(Q_i | D_i) \quad (4)$$

Also applies to  $L^w$  and if (2) defined with  $R = (Q|D) = \frac{(\gamma R)}{(\sum_{i=1}^n \exp(\gamma R))^\alpha}$  then substitution to (1) and (4) is obtained  $L'_1$  and  $L'_2$ .

The following is loss function formula for generator:

$$L_G = - \frac{1}{2} [E_{z \sim p_z} \log D(x)] + E_{z \sim p_z} \log [D(x, \varphi(z))] \quad (5)$$

When the input is a noise distribution, many probabilities are generated due to vector differences. Some exist mode is preferred over other modes. The mode-seeking loss function tries to amplify the distribution of minor modes between major modes. When the noise distribution is used to generate two different samples based on the same input. The first point  $z_1$  causes the formation of a mode image  $I_1 = G(z_1)$  and the second point  $z_2$  points to the image  $I_2 = G(z_2)$  after applying generate image and text ( $c$ ).  $D(\dots)$  is a distance function that calculates the distance between two tensors or vectors. The numerator calculates the distance between object vectors while the denominator is the distance between ordinary vectors. The purpose of this formula is to maximize the distance between vector objects. The following is the MSGAN loss function formula.

$$L_{MSG} = \max_G \left( \frac{D(G(z_1), G(z_2))}{D(z_1, z_2)} \right) \quad (6)$$

Based on (1) until (6). The calculation of the loss function formula is (7)

$$L = \sum_i L_G + L_{DAMSM} + L_{MSG} \quad (7)$$

The structure of the GAN architecture that adopts DCGAN [10], the difference between the input generators is the distribution of noise and text. The text encoder does not use word embedding but a Skip-thought model[22]. Adoption of the DCGAN structure as it mimics the workings of discriminator networks studying the appropriate relationship between description of text and image content with GAN-CLS and GAN-INT function.

At the start of the training, the discriminator ignores the text and easily reject the image generated from the generator, because it doesn't look reasonable, but after the generator receives feedback update weight from the discriminator, the generator learns to generate a reasonable generated image and learns to align it with the text, the the discriminator also learn to evaluate the resulting image to satisfy this conditioning problem. Based on this intuition it may complicate the dynamics of learning, but GAN-CLS pays attention to this by adding an alternate decision to the discriminator, namely a third input consisting of the original image with mismatched text. It is in accordance with the DAMSM loss function.

GAN-INT works with word embedding on the generator side. GAN-INT utilizes manifold interpolation learning to generate additional text for word embedding during training. This matter serves to make the generator stronger to generate images and can deceive the discriminators from being able to distinguish the output. The results of the added text to be inserted do not need to match the facts so that GAN-INT does not require a cost.

## VI. CONCLUSION

Word embedding glove and word2vec are static word models which are the most widely used word embedding. The availability of pretrained word embedding models and with vocabulary area coverage can reduce the high computational requirements for training word embedding models. In this work, we explore a simple technique, namely concatenate word embedding, which contributes to developing GAN for the task of generating images from text and is expected to improve performance on a more robust GAN. Based on the explanation of our research method, this model can be implemented and can be developed for various GAN tasks such as style transfer, image to image, face inpainting or image repair semantically, and super resolution.

## ACKNOWLEDGMENT

Thank to Indonesia Endowment Fund for Education (LPDP) Ministry of Finance Republic Indonesia have been made possible this research and presentation.

## REFERENCES

- [1] Martin C Nwudigwu, "Neural Networks, Artificial Intelligence and the Computational Brain," *ArXiv*, 2021.
- [2] Ian J. Goodfellow *et al.*, "Generative Adversarial Nets," 2014.
- [3] Y. Jing, Y. Yang, Z. Feng, J. Ye, and M. Song, "Neural style transfer," *IEEE Trans Vis Comput Graph*, 2019.
- [4] L. A. Gays, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2016, pp. 2414–2423.

- [5] J.-Y. Z. T. Z. A. A. E. P. Isola, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2016, pp. 1125–1134.
- [6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.
- [7] T. A. S. L. J. L. T. Karras, "Progressive growing of gans for improved quality, stability, and variation," 2018.
- [8] J. Yu, Z. Liu, J. Yang, X. Shi, X. Lu, and T. S. Huang, "Free-form image inpainting with gated convolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 4471–4480.
- [9] R. A. Yeh, C. Chen, T.-Y. Liu, A. G. Schwing, and M. N. M. Hasegawa-Johnson, "Semantic image inpainting with deep generative models," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2016, pp. 5485–5493.
- [10] Scott Reed and Zeynep Akata and Xiaohu Yan and Lajunqun Logezwaran and Bernd Schiele and Honglak Lee, "Generative Adversarial Text-to-Image Synthesis," *Proc. Mach Learn Res*, 2016.
- [11] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2016, pp. 4681–4690.
- [12] J. Donahue and K. Simonyan, "Large scale adversarial representation learning," *Advances in Neural Information Processing Systems*, pp. 10452–10552, 2019.
- [13] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," 2013.
- [14] R. S. and C. D. M. J. Pennington, "GloVe: Global Vectors for Word Representation," 2014.
- [15] S. K. Gorti and J. Ma, "Text-to-Image-to-Text Translation using Cycle Consistent Adversarial Networks," Aug. 2018, [Online]. Available <http://arxiv.org/abs/1808.04531>
- [16] Han Zhang and Tao Xu and Hongsheng Li and Shouping Zhang and Niogang Wang and Xiaohu Huang and Demin Metzias, "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks," *International Conference on Computer Vision (ICCV)*, 2017.
- [17] Sai Rajeswar, Sandeep Subramanian, Francis Dutil, Christopher Joseph Pal, and Aaron C. Coveville, "Adversarial generation of natural language," *CoRR*, 2017.
- [18] Rezaul Naderalvejal and Ebru Akcayin Sener, "Sentiment aware word embeddings using refinement and semi-contextualized learning approach," *Neurocomputing*, vol. 405, pp. 149–180, 2020.
- [19] Seyad Mahdi Razaeeinia, Rezaollah Bahrami, Ali Ghodabi, and Hadi Veisai, "Sentiment analysis based on improved pre-trained word embeddings," *Expert Syst. Appl.*, vol. 117, pp. 139–147, 2019.
- [20] A. D. Perillo, Lucchi, and T. Hofmann, "Faces à la Carte: Text-to-Face Generation via Attribute Disentanglement," *European Conference on Computer Vision*, pp. 482–499, 2020.
- [21] Tao Xu and Pengchuan Zhang and Qiyuan Huang and Han Zhang and Zhu Guo and Xiaohu Huang and Xiaodong He, "AttriGAN: Fine-Grained Text-to-Image Generation with Attentional Generative Adversarial Networks," *CoRR*, 2017.
- [22] Ryan Kiros et al., "Skip-Thought Vectors," *CoRR*, 2015.

# Cek Plagiasi 12

---

ORIGINALITY REPORT

---

**23%**

SIMILARITY INDEX

**16%**

INTERNET SOURCES

**16%**

PUBLICATIONS

**7%**

STUDENT PAPERS

---

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

---

3%

★ Submitted to University of Melbourne

Student Paper

---

Exclude quotes      On

Exclude matches      Off

Exclude bibliography      On