USULAN PENELITIAN

VIDEO UNDERSTANDING PELANGGARAN LALU LINTAS BERDASARKAN CITRA *CONTEXT* BAHASA INDONESIA MENGGUNAKAN *MULTI TASK DEEP LEARNING*

KOMANG AYU TRIANA INDAH

FAKULTAS TEKNIK
UNIVERSITAS UDAYANA
DENPASAR
2021

USULAN PENELITIAN

VIDEO UNDERSTANDING PELANGGARAN LALU LINTAS BERDASARKAN CITRA *CONTEXT* BAHASA INDONESIA MENGGUNAKAN *MULTI TASK DEEP LEARNING*

KOMANG AYU TRIANA INDAH
NIM. 2091011009

VI

PROGRAM DOKTOR
PROGRAM STUDI DOKTOR ILMU TEKNIK
FAKULTAS TEKNIK
UNIVERSITAS UDAYANA
DENPASAR
2021

Lembar Persetujuan Pembimbing Akademik

PRAPROPOSAL PENELITIAN DISERTASI INI TELAH DISETUJUI

PADA TANGGAL 17 Mei 2021

Pembimbing Akademik

(Prof. Dr. Ir. Made Sudarma, M.A.Sc., IPU)
NIP. 196512311993031189

Mengetahui:

Koordinator Program Studi Doktor Ilmu Teknik
Fakultas Teknik Universitas Udayana

(Prof. Ir. I Nyoman Arya Thanaya, ME, Ph.D.)
NIP. 196011081988031002

Praprosal Penelitian Disertasi Ini Telah Disetujui dan Dinilai
oleh Panitia Penguji pada
Program Doktor Ilmu Teknik
Program Pascasarjana Universitas Udayana
Pada tanggal ……..

Panitia Penguji Praproposal Penelitian Disertasi :

Ketua : Prof. Dr. Ir. Made Sudarma, M.A.Sc., IPU (PA)

Anggota :

1. Prof. Dr. I Ketut Gede Darma Putra, S.Kom., M.T.
2. Prof. Ir. Rukmi Sari Hartati, M.T.,Ph.D
3. Dr.Eng.I Putu Agung Bayupati, S.T.,M.T.
4. Dr.A.A Kompiang Oka Sudana, S.Kom., M.T.
5. Dr. Ida Bagus Gede Manuaba, ST., M.T.
6. Dr. I Made Sukarsa, S.T.,M.T.

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING AKADEMIK iii

PANITIA PENGUJI PRAPROPOSAL PENELITIAN DISERTASI iv

DAFTAR ISI v

DAFTAR GAMBAR vii

DAFTAR TABEL viii

DAFTAR LAMPIRAN ix

BAB I PENDAHULUAN

1.1 Latar Belakang 1

1.2 Perumusan Masalah 5

1.3 Tujuan Penelitian 6

1.4 Manfaat Penelitian 7

1.5 Batasan Masalah 8

1.6 Keaslian Penelitian (Novelty) 9

BAB II TINJAUAN PUSTAKA

2.1 *State Of The Art* 11

2.2 Teori dan Metode 35

2.2.1 *Computer Vision* 35

2.2.2 *Multi Task Learning* 37

2.2.3 *Recurrent Neural Network* 40

2.2.4 *Bi LSTM Model* 45

BAB III KERANGKA BERFIKIR, KOMNSEP PENELITIAN DAN HIPOTESIS

3.1 Kerangka Berfikir 47

3.2 Konsep Penelitian 50

BAB IV METODE PENELITIAN

4.1 Peta Penelitian 51

4.2 Rancangan Penelitian 51

4.3 Teknik Pengujian 62

4.4 Lokasi dan Waktu Penelitian 63

4.5 Instrumen Penelitian..... 63

4.6 Prosedur Penelitian..... 64

DAFTAR PUSTAKA 66

DAFTAR GAMBAR

Gambar 2.1	*Fishbone* Penelitian	33
Gambar 2.2	Tree Diagram Penelitian	34
Gambar 2.3	Proses Pada *Computer Vision*	37
Gambar 2.4	*Hard Parameter Sharing for Multi-Task Learning*	38
Gambar 2.5	*Soft Parameter Sharing for Multi-Task Learning*	39
Gambar 2.6	*Deep Relationship Network* dengan Lapisan Konvolusional Bersama dan Sepenuhnya Terhubung Dengan Prior Matriks ..	40
Gambar 2.7	Image Completions Sampled From a Pixel RNN	41
Gambar 2.8	Proses Context Citra	43
Gambar 2.9	Diagonal BiLSTM	46
Gambar 3.1	Konsep Model Text Sebagai Ruang Label yang Menghasilkan Deskripsi pada Area Gambar	48
Gambar 3.2	Model Kumpulan Gambar dan Deskripsi yang Sesuai	49
Gambar 3.3	Konsep Penelitian	50
Gambar 4.1	Peta Penelitian	52
Gambar 4.2	Simulasi Pembuatan Grafik Adegan dan Pembuatan Teks Gambar	53
Gambar 4.3	Kerangka Keseluruhan dari Model Citra Understanding	54
Gambar 4.4	Jaringan Deteksi Objek dan Wilayah Teks	54
Gambar 4.5	Alur Keseluruhan dari Model Visual Understanding	55
Gambar 4.6	Eksplorasi Data-Set	55
Gambar 4.7	Diagram Evaluasi Citra-Kalimat	60
Gambar 4.8	Training Workflow	62
Gambar 4.9	Siklus Hidup Pengembangan Sistem	65

DAFTAR TABEL

Tabel 2.1 Rangkuman Penelitian Sebelumnya tentang Multi *Task Deep Learning* 11

Tabel 2.2 Rangkuman Penelitian Sebelumnya tentang *Computer Vision, (CNN), (RNN), (LSTM), (CCN) dan (RCN)* 13

Tabel 2.3 Penelitian Sebelumnya Tentang Deteksi Pelanggaran Lalu Lintas 15

Tabel 2.4 *Matrix State Of The Art* 16

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pemahaman akan sebuah citra (gambar dan video) atau *visual understanding* merupakan salah satu inti dari konsep *computer vision* yang diteliti secara ekstensif. *Computer vision* merupakan proses yang mengintegrasikan sejumlah besar proses untuk persepsi visual, seperti akuisisi citra, pengolahan citra, pengenalan dan membuat keputusan. *Visual understanding* adalah salah satu elemen inti dari konsep *computer vision* yang terdiri dari klasifikasi citra, deteksi objek dan segmentasi akan dilakukan melalui metode *multi task deep learning* dengan metode CNN (Convolution Neural Network) dengan teknik *Recurrent Neural Network* (RNN) *dan Long Term Short Memory* (LSTM). Metode *multi task deep learning* pada *video understanding* diimplementasikan pada proses konversi *image to text* (*context*) melalui proses *multi layer* untuk mengekspresikan pemahaman gambar yang kompleks. Kebutuhan akan interpretasi sebuah citra visual kedalam bentuk teks kalimat dapat diimplementasikan ke berbagai bidang, diantaranya di bidang kedokteran, informasi geografis, deteksi pelanggaran lalu lintas, ramalan cuaca dan sebagainya (Guo *et al.*, 2016). Bagaimana menterjemahkan sebuah visualisasi video kedalam bentuk semantik/kalimat kompleks yang berbahasa Indonesia merupakan objek dari penelitian ini yang meliputi citra klasifikasi, deteksi objek, dan segmentasi semantik (Shin and Kim, 2018). Pada penelitian sebelumnya berfokus pada informasi yang kurang detail

karena hanya mengidentifikasi objek yang termasuk dalam gambar dan lokasi objek ke dalam bentuk keterangan/*content labelling* yang diinputkan secara manual serta serangkaian kategori visual yang sifatnya tetap. (Kinghorn, Zhang and Shao, 2018) Salah satu masalahnya yaitu bagaimana menjawab pertanyaan visual yang merujuk pemahaman ekspresi yang menjelaskan detail penampakan visual secara kompleks. Dimana konsep tersebut sangat terbatas jika dibandingkan sejumlah variasi deskripsi yang dibuat oleh manusia, baik dalam hal semantik maupun kosakata dan tata bahasa terutama Bahasa Indonesia.

Untuk meningkatkan pengawasan terhadap arus lalu lintas dan kedisiplinan pengguna jalan dalam mematuhi aturan, maka pada bulan Maret 2021 di beberapa kota besar di Indonesia diberlakukan tilang elektronik atau ETLE (*Electronic Traffic Law Enforcement*), yang merupakan fasilitas bidang lalu lintas dan penegakan hukum berbasis *Information Technology* (IT), (Z.Zang, 2010) yang dilansir dari halaman NTMC (*National Traffic management Center*) POLRI pada tanggal 1 Maret 2021. Video rekaman pelanggaran lalu lintas di beberapa ruas jalan melalui kamera CCTV menggunakan teknologi ETLE (*Electronic Traffic Law Enforcement*). Pada tahap awal peluncuran ETLE secara nasional akan diterapkan di tiga Polda yakni di Polda Jawa Barat, Polda Jawa Tengah, dan Polda Riau serta empat Polresta, yakni Polresta Jambi, Polresta Gresik, Polresta Batam dan Polresta Padang. Perekaman bukti pelanggaran dilakukan melalui kamera CCTV dan *mobille camera* yang terpasang dibeberapa titik pertigaan dan perempatan ruas jalan. (K. Klubsuwan, 2013) .

Permasalahan dari sistem ini adalah terdapat ratusan bahkan ribuan video yang terekam kamera setiap harinya dan diperlukan pengamatan yang seksama oleh petugas untuk menentukan jenis pelanggaran yang nantinya terkait dengan penentuan sanksi pelanggaran. Oleh karena itu diperlukan sebuah sistem interpreter objek citra video untuk merepresentasikan video rekaman pelanggaran lalu lintas (*traffic violence*) oleh pengendara kendaraan bermotor kedalam bentuk teks/kalimat (Z.Zang, 2010) ke dalam Bahasa Indonesia. Tindakan pemberlakuan sanksi pelanggaran kepada masyarakat selama ini dilakukan langsung oleh petugas kepolisian maupun melalui analisa rekaman kamera ETLE (*Electronic Traffic Law Inforcement*) yang terpantau dari ATCS (*Area Traffic Control System*), dengan mengambil detail data-data pelanggaran lalu lintas berdasarkan TNKB (Tanda Nomor Kendaraan), fitur pelanggaran rambu, marka jalan, dan *traffic light* yang kemudian disimpan pada database *back office* ETLE (*Electronic Traffic Law Inforcement*) di RTMC *(Road Traffic Management Corporation)* Polda setempat. Nantinya petugas akan mengidentifikasi data kendaraan menggunakan *Electronic Registration and Identification* (ERI) sebagai sumber data kendaraan. Proses yang dilakukan oleh sistem ETLE hanya mengidentifikasi pelaku pelanggaran, dan berdasarkan data kendaraan, namun untuk jenis pelanggarannya masih dianalisa oleh petugas melalui video rekaman yang ada pada database. Berdasarkan data pelanggaran lalu lintas yang terpantau melalui ATCS (*Area Traffic Control System*) data yang dianalisa pada sistem ETLE di RTMC dan Dinas Perhubungan harus mengklasifikasikan pelanggaran lalu lintas menjadi beberapa komponen, diantaranya (Aini, Hutapea and Ramadhanie, 2020):

- Jenis Pelanggaran Menggunakan TNKB (Tanda Nomor Kendaraan Bermotor) palsu
- Pengendara Tidak Menggunakan Helm (Kendaraan Roda Dua)
- Pengendara tidak Memakai Sabuk Pengaman (Kendaraan Roda Empat)
- Pengendara Berhenti di Zebra Cross
- Pengendara Berhenti di RHK (Ruang Henti Khusus)
- Pengendara Berhenti Melebihi Stopline
- Pengendara Merokok
- Pengendara Kelebihan Muatan (Kendaraan Roda Empat)
- Pengendara Berboncengan Tiga (Kendaraan Roda Dua)
- Pelanggaran Rambu lalu lintas dan Marka Jalan
- Pelanggaran Melawan Arus
- Pelanggaran Pembatasan Jenis Kendaraan Tertentu pada beberapa ruas jalan

Dari jenis pelanggaran diatas yang terpantau oleh ATCS melalui kamera CCTV ataupun *mobile camera* pada lokasi simpang jalan yang terdapat *traffic light*, dibutuhkan pengamatan yang seksama dari petugas untuk menentukan jenis pelanggarannya, yang terkait dengan pemberlakuan sanksi. Melalui deskripsi text dari sebuah model video dalam bentuk frame akan didapatkan sebuah konsep korespondensi antara bahasa dan visual ke dalam bentuk teks/kalimat. Sinkronisasi ini didasarkan pada teknik kombinasi baru dengan model *Context-based Captioning and Scene Graph Generation Network (C2SGNet)* yang merupakan metode *Multitask Deep Learning* yang merupakan kombinasi dari

sistem *Convolutional Neural Network* (CNN) pada area gambar, *Recurrent Neural Network* (RNN) dengan LSTM (*Long Short Term Memory*) pada area kalimat, CCN (*Caption Content Network*) dan RCN (*Relationship Context Network*) pada proses perlabelan dan relasi antar objek serta adegan beserta tujuan terstruktur yang menyelaraskan dua modalitas melalui penyematan multimodal (Bai *et al.*, 2018). RNN masuk dalam kategori *deep learning* karena data diproses melalui banyak lapis (*layer*). RNN telah mengalami kemajuan yang pesat dan telah merevolusi bidang-bidang seperti pemrosesan bahasa alami atau *Natural Language Processing* (NLP), pengenalan suara, sintesa musik, pemrosesan data finansial seri waktu, analisa deret DNA, analisa video, dan sebagainya.

Pada penelitian ini, akan dibangun sebuah sistem deteksi karakteristik citra video melalui pemahaman (*video understanding*) pada pelanggaran lalu lintas untuk menghasilkan keterangan gambar pada setiap frame beserta ekspresi gambar tingkat tinggi berupa teks dan grafik adegan melalui teknik klasifikasi citra untuk ditransformasi dalam bentuk teks berbahasa Indonesia, sehinga memudahkan dalam mengidentifikasikan jenis pelanggaran yang ada kaitannya dengan penentuan sanksi oleh petugas sesuai undang-undang yang berlaku. Objek penelitian difokuskan untuk video pelanggaran lalu lintas perkotaan (urban traffic), dengan mengumpulkan dataset objek terkait dari beberapa titik kamera CCTV dan *mobile camera* di jalur simpang perempatan dan pertigaan jalan.

1.2 Rumusan Masalah

Pemahaman visual (*video understanding*) merupakan salah satu elemen dari *computer vision* yang meliputi citra klasifikasi, deteksi objek dan segmentasi semantik. Penelitian sebelumnya hanya berfokus pada informasi umum untuk mengidentifikasi objek beserta lokasinya, tetapi tidak menyelesaikan pemahaman konten gambar yang kompleks, untuk itu pada penelitian ini dirumuskan permasalahan yaitu :

1. Bagaimana merepresentasikan transformasi pemahaman objek video ke dalam conteks kalimat berbahasa Indonesia secara semantik dan grafik adegan dalam hubungan antar objek secara umum yang terdiri dari *object recognition, object detection, semantic segmentation* dan *image captioning* pada objek video pelanggaran lalu lintas yang terekam kamera CCTV.

   Rumusan permasalah diatas diuraikan menjadi sub permasalahan, antara lain:

   a. Bagaimana menghasilkan pemahaman tingkat tinggi sebuah visual citra gambar dan grafik adegan secara bersamaan dalam bentuk konversi ke dalam bentuk teks menggunakan model *Context-based Captioning and Scene Graph Generation Network (C2SGNet).*
   b. Bagaimana merancang model deteksi objek dan deteksi hubungan dari sebuah gambar masukan menjadi kalimat dengan Bahasa Indonesia berdasarkan pendekatan CCN (*Caption Content Network*) dan RCN (*Relationship Context Network*) dalam mekanisme *Multi Task Deep Learning.*
   c. Bagaimana mengimplementasikan metode *Convolutional Neural Network* dalam mengekstraksi objek dalam gambar.

d. Bagaimana menghasilkan kalimat/caption yang sesuai dengan fitur gambar dengan menggunakan metode RNN (Recurrent Neural *Network*) dan LSTM (*Long Short-Term Memory*) dalam struktur bahasa yang kompleks dengan dataset kosakata Bahasa Indonesia.
e. Bagaimana melakukan evaluasi kinerja hasil penelitian dengan menggunakan metode *Visual Dataset Genom*, yang merupakan kumpulan data dan basis pengetahuan untuk menghubungkan konsep gambar terstruktur ke dalam bantuk bahasa.

1.3 Tujuan

Tujuan umum dari penelitian ini adalah merepresentasikan gambar masukan dengan konsep *image understanding* menjadi kalimat Bahasa Indonesia secara semantik dan grafik adegan beserta karakteristik pelengkapnya, pada objek video pelanggaran lalu lintas, yang diinterprestasikan berdasarkan durasi dan frame.

Adapun tujuan khusus dari penelitian ini adalah:

a. Menghasilkan ekstraksi citra video kedalam frame citra gambar dengan menggunakan teknik *Context-based Captioning and Scene Graph Generation Network (C2SGNet*) dalam metode *Multi Task Deep Learning*?
b. Merancang model deteksi objek dan deteksi hubungan dari sebuah gambar masukan menjadi kalimat dengan bahasa Indonesia berdasarkan pendekatan CCN (*Caption Content Network*) dsn RCN (*Relationship Context Network*).
c. Mengimplementasikan metode *Convolutional Neural Network* dalam mengekstraksi dan klasifikasi objek dalam gambar.

d. Menghasilkan kalimat/caption yang sesuai dengan fitur gambar dengan menggunakan metode RNN (*Recurrent Neural Network*) dan LSTM (*Long Short-Term Memory*).

e. Bagaimana membuat *evaluation and trained model* dengan menggunakan Visual Dataset Genom untuk pengujian dan analisa sistem.

1.4 Manfaat Penelitian

Adapun manfaat yang diperoleh melalui penelitian ini antara lain manfaat teoritis dan manfaat praktis:

a. Manfaat secara teoritis pada penelitian ini diharapkan menghasilkan:

- Pengembangan metode *Deep Learning* dan *Image Processing* melalui sebuah sistem yang mampu memecahkan masalah pemahaman citra tingkat tinggi (*image understanding*) untuk konversi citra visual dalam context kalimat dari sebuah objek dalam 3 lapisan yaitu deteksi objek, deteksi hubungan dan pembuatan teks dalam Bahasa Indonesia. Objek penelitian menggunakan dataset *urban traffic* dalam mendeteksi pelanggaran lalu lintas di suatu objek jalan perkotaan dalam kalimat sematik berbahasa Indonesia.
- Bagi pengembangan teknologi *machine learning* dengan *multi task deep learning* dengan penggunaan beberapa metode diantaranya:

  1. *Context-based Captioning and Scene Graph Generation Network (C2SGNet*) yang merupakan model jaringan saraf terintegrasi dalam sebuah sistem pemahaman citra visual yang menghasilkan natural teks

bahasa dan grafik pemandangan (Subjek Predikat Objek) untuk gambar input menggunakan informasi konteks.

2. *Convolutional Neural Network* (CNN) berfungsi untuk mengekstraksi fitur visual berupa gambar, melalui *Recurrent Neural Network* (RNN) terutama tipe LSTM untuk pembuatan caption berbagai wilayah gambar dari elemen objek dengan mendeteksi hubungan antar objek.

b. Manfaat secara praktis bagi masyarakat, penelitian memberikan kontribusi dalam mengembangkan sistem dengan:

- Hasil penelitian dapat dijadikan acuan pemberian data informasi dari jenis pelanggaran lalu lintas dan berbagai fenomena citra visual, sehingga memudahkan dalam analisa objek lebih efektif terutama dalam pengambilan keputusan jenis pelanggaran lalu lintas oleh petugas ATCS (*Area Traffic Control System*) yang menganalisa data pada sistem data Dinas Perhubungan setempat.
- Sistem video understanding dapat diimplementasikan sebagai visual interpreter untuk memberikan informasi berdasarkan subjek, objek dan relasi diantaranya dalam bentuk teks/kalimat.

1.5 Batasan Masalah

Pada penelitian ini, sistem dibatasi pada :

1. Konversi video dalam context kalimat berasal dari objek data rekaman CCTV melalui ETLE (*Electronic Traffic Law Enforcement*) yang berupa

pelanggaran lalu lintas yang melalui 3 lapisan proses yaitu deteksi objek, deteksi hubungan dan pembuatan teks ke dalam Bahasa Indonesia.

2. Klasifikasi context data menggunakan metode *Recurrent Neural Network* dengan tipe LSTM (*Long Short Term Memory*) *Bidirectional*.
3. Penelitian tidak dilakukan untuk menentukan jenis-jenis sanksi pelanggaran, tapi fokus dalam proses identifikasi, ekstraksi dan segmentasi objek ke dalam bahasa semantik pada video pelanggaran lalu lintas.
4. Dataset yang dipergunakan dibagi menjadi 3 subset, yaitu Data Training, Data *Validation* dan Data *Test*, dengan pembagian 70%-80% *Training*, 10%-20% Validasi dan 10% *Test*, dimana antar subset tidak dibuat *overlap* karena akan merusak proses training model.
5. Sistem diimplementasikan dengan menggunakan Bahasa Pemrograman Phyton dan dibangun berbasis web.

1.6 Keaslian Penelitian (Novelty)

Kebaharuan dari penelitian ini yaitu:

- Pada penelitian sebelumnya konsep pemahaman visual/image (*image understanding*) berdasarkan context citra menjadi kalimat dengan pola semantik pada umumnya menggunakan teks kalimat berbahasa Inggris, sejauh yang diketahui penulis dari penelitian-penelitian yang telah dibahas sebelumnya, belum ada penelitian yang membangun sistem context berbahasa Indonesia. (Bouwmans *et al.*, 2015), (Jeeva and Sivabalakrishnan, 2015), (Guo *et al.*, 2016), (Wiriyathammabhum *et al.*,

2017), (Shin and Kim, 2018), (Jácome-Galarza *et al.*, 2020), (Kinghorn, Zhang and Shao, 2018b), (Bai and An, 2018), (Wang *et al.*, 2019), (Sarraf, Azhdari and Sarraf, 2021). (Liu *et al.*, 2019) (Staniute and Šešok, 2019).

- Pada penelitian sebelumnya sejauh yang diketahui penulis belum ada metode *video understanding* dengan *Multi Task Deep Learning* serta model *Context-based Captioning and Scene Graph Generation Network (C2SGNet)* untuk objek data video pelanggaran lalu lintas, penelitian terhadap objek ini hanya berupa deteksi dan monitoring pelanggaran lalu lintas dengan metode *Deep Learning* serta kajian pustaka. (Ariyoga, Rahmadi and Rajagede, 2021), (Z.Zang, 2010), (K. Klubsuwan, 2013) (Singh, Unadkat and Kanani, 2019) (Alkan *et al.*, 2019).

BAB II

TINJAUAN PUSTAKA

2.1 State Of The Art

Pembangunan sistem representasi citra visual dalam context Bahasa Indonesia dari sebuah objek dalam 3 lapisan yaitu deteksi objek, deteksi hubungan dan pembuatan teks pada penelitian ini didasarkan pada penelitian-penelitian yang telah dilakukan sebelumnya, seperti yang dirangkum pada Tabel 2.1, 2.2, 2.3 dan 2.4 berikut ini.

Tabel 2.1
Rangkuman Penelitian Sebelumnya tentang *Multi Task Deep Learning*

No.	Penulis Dan Tahun	Metode	Judul Penelitian
1	Qiu, Junfei Wu, QihuiDing, Guoru Xu, Yuhua Feng, Shuo, 2016	*Machine Learning*	A survey of machine learning for big data processing
2	Doersch, Carl Zisserman, Andrew, 2017	Multi-task Self-Supervised Visual Learning	Multi-task Self-Supervised Visual Learning
3	Chen, Bor Chun Ghosh, Pallabi Morariu, Vlad I.Davis, Larry S, 2017	*Multi Task Deep Learning*	Detection of Metadata Tampering Through Discrepancy between Image Content and Metadata Using Multi-task Deep Learning
4	Ruder, Sebastian, 2017	*Multi Task Learning*	An Overview of Multi-Task Learning in Deep Neural Networks

No.	Penulis Dan Tahun	Metode	Judul Penelitian
5	Cipolla, Roberto Gal, Yarin Kendall, Alex, 2018	*Multi Task Learning*	Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics
6	Pugaliya, Hemant Saxena, Karan Garg, Shefali Shalini, Sheetal Gupta, Prashant Nyberg, Eric Mitamura, Teruko, 2019.	*Multi Task Learning*	Multi-task learning for filtering and re-ranking answers using language inference and question entailment
7	Hessel, Matteo Soyer, Hubert Espeholt, Lasse Czarnecki, Wojciech Schmitt, Simon van Hasselt, Hado, 2019	*Multi-task deep reinforcement learning*	*Multi-task deep reinforcement learning with PopArt*
	Ciaparrone, Gioele Luque Sánchez, Francisco Tabik, Siham Troiano, Luigi Tagliaferri, Roberto Herrera, Francisco, 2020	*Deep Learning*	Deep learning in video multi-object tracking: A survey
8	Liu, Xiaodong He, Pengcheng Chen, Weizhu Gao, Jianfeng, 2020	Multi-task deep neural networks	Multi-task deep neural networks for natural language understanding

Tabel 2.2
Rangkuman Penelitian Sebelumnya tentang *Computer Vision, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Caption Content Network (CCN) dan Relationship Context Network (RCN)*

No.	Penulis Dan Tahun	Metode	Judul Penelitian
1	Brutzer, S., Hoferlin, B., Heidemann, G, 2011	CVPR (Computer Vision and Pattern Recognition)	Evaluation of background subtraction techniques for video surveillance “,Computer Vision and Pattern Recognition (CVPR)
2	Kelvin Xu Jimmy Lei Ba Ryan Kiros Kyunghyun Cho Aaron Courville Ruslan Salakhutdinov Richard S. Zemel Yoshua Bengio, 2016	R-CNN	Neural Image Caption Generation With Visual Attention
3	Philip Kinghon, L.Zhang,L.Shao,2016	Recurrent Neural Network (RNN)	A Region-Based Image Caption Generator with Refined Description
4	Guo, Yanming Liu, Yu Oerlemans, Ard Lao, Songyang Wu, Song Lew, Michael S, 2016	Deep Learning Algorithm	Deep Learning for Visual Understanding:A Review
5	Justin Johnson, Andrej Karpathy, Li Fei-Fei, 2016	Convolutional Localization Network	Fully Convolutional Localization Networks for Dense Captioning
6	Peratham Wiriyathammabhum, Douglas, Summer-Stay, Cornelia Fermuller, Yiannis Aloimonos, 2016	Computer Vision dan NLP (Natural Language Processing)	Computer Vision and Natural Language Processing Apprioaches in Multimedia and Robotics
7	Tjatur Kandaga Gautama, Antonius Hendrik, Riskadewi, 2016	Computer Vision Algoritma Sift	Pengenalan Objek pada Computer Vision dengan Pencocokan Fitur Menggunakan Algoritma

No.	Penulis Dan Tahun	Metode	Judul Penelitian
			SIFT Studi Kasus: Deteksi Penyakit Kulit Sederhana.
8	Somak Aditya, Yezhou Yang, Chitta Baral, 2017	SDG (Scene Description Graph)	Image Understanding using Vision and Reasoning Through Scene Description Graph
9	Dewi, Syarifah Rosita, 2018	Tensorflow Dan Convolutional Neural Network	Deep Learning Object Detection Pada Video Menggunakan Tensorflow Dan Convolutional Neural Network
10	Donghyeop Shin, Incheol Kim, 2018	Faster R-CNN (Convolutional Neural Network)	Deep Image Understanding Using Multilayer Context
11	Shuang Bai, Shan An, 2018	Deep Neural Network	A Survey an Automatic Image Caption Generation
12	Ghosh, Swarnendu Das Nibaran, Das Ishita, Maulik, Ujjwal, 2019	Deep Learning Convolutional Neural Network	Understand Deep Learning Technique For Image Segmentation
13	Raimonda Staniute, Dmitrij Sesok, 2019	LSTM (Long Short Term Memory) dan SLR (Systematic Literature Review)	A Systematic Literature Review on Image Captioning
14	Jácome-Galarza, Luis Roberto Realpe-Robalino, Miguel Andrés Chamba-Eras, Luis Antonio Viñán-Ludeña, Marlon Santiago, 2020	Computer Vision	Computer Vision for Image Understanding A Comperensive Review

No.	Penulis Dan Tahun	Metode	Judul Penelitian
15	Haneol Jang, Jong -Uk Hou,2020	Detecting Contextual Abnormality Using Convolutional Neural Network	Exposing Digital Image Forgeries by Detecting Contextual Abnormality Using Convolutional Neural Networks

Tabel 2.3
Penelitian Sebelumnya Tentang Deteksi Pelanggaran Lalu Lintas dengan *Deep Learning*

No.	Penulis dan Tahun	Metode	Judul Penelitian
1	Z.Zang, 2010	*Object recognition* dengan Hopfield's Neural Model	Research on the taxi traffic accident and violation identification model
2	K. Klubsuwan, W. Koodtalang and S. Mungsing, 2013	Mean Square Displacement (MSD)	Traffic Violation Detection Using Multiple Trajectories Evaluation of Vehicles
3	Singh, Vedant Unadkat, Vyom Kanani, Pratik, 2019	*Deep learning* menggunakan YOLOv3 dan model ROI (*Region of Interest*)	Intelligent traffic management system
4	Ambekar, Shubham Dagade, Mangesh, 2019	*Convolutional Neural Network*	Traffic signal violation monitoring through video surveillance using CNN
5	Z.Aini, F.Hutapea, N.Ramadhanie, 2020	Deteksi objek melaui perangkat CCTV	Implementasi Sistem Pegawas CCTV Di Kota Tanjungpinang (Studi Kasus Dinas Perhubungan)
6	D.Ariyoga, R.Rahmadi, R.Rajagede, 2021	*Deep Learning*	Penelitian Terkini Tentang Sistem Pendeteksi Pelanggaran Lalu Lintas Berbasis Deep Learning: Sebuah Kajian)

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
1	Brutzer, S., Hoferlin, B., Heidemann, G, 2015	*Evaluation of background subtraction techniques for video surveillance ",Computer Vision and Pattern Recognition (CVPR)*	-Detail dari pemahaman urutan rangkaian adegan video, dan *multiple image* dari latar yang sama. -Perspektif untuk memberikan studi nomenklatur dari langkah-langkah pemrosesan umum yang digunakan dalam algoritme deteksi perubahan yang digunakan untuk pengawasan video waktu nyata (*video surveillace*). Model ini mencakup teknik pemodelan prediktif dan latar belakang	-Metode Vibe dapat mengklasifikasi model latar belakang berdasarkan instrumen matematika yang digunakan. Dari hasil perbandingan kecepatan pemrosesan dengan metode waktu nyata. Dapat disimpulkan bahwa lebih efisien dan banyak frame dapat diproses per detik -Algoritme Self-Organizing Background Subtraction (SOBS) secara akurat menangani adegan yang berisi gerakan latar belakang, variasi iluminasi bertahap, dan bayangan yang ditimbulkan oleh objek bergerak, dan kuat terhadap deteksi palsu untuk berbagai jenis gambar yang diambil dengan kamera stasioner.	Relevansinya yaitu metode yang digunakan menjadi referensi dalam deteksi perubahan beberapa gambar dari pemandangan yang sama secara seketika. Tugas ini diperlukan karena membawa sejumlah besar aplikasi bidang subjek yang beragam Perspektif utama adalah untuk memberikan studi nomenklatur dari langkah-langkah pemrosesan umum dan aturan keputusan utama yang digunakan dalam algoritme deteksi perubahan tingkat lanjut pada *real time video surveillance.*
2	Kelvin Xu Jimmy Lei Ba Ryan Kiros Kyunghyun Cho Aaron Courville Ruslan Salakhutdinov Richard S. Zemel Yoshua Bengio, 2016	*Neural Image Caption Generation With Visual Attention*	-Model berbasis *attention* untuk mendeskripsikan konten gambar dengan melatih model ini secara deterministik menggunakan teknik propagasi mundur standar -Visualisasi dilakukan untuk memperbaiki pandangannya pada objek yang dominan sambil menghasilkan kata-kata yang sesuai dalam urutan keluaran.	-Meningkatkan kualitas pembuatan teks secara signifikan menggunakan kombinasi jaringan saraf konvolusional (konvnet) untuk mendapatkan representasi vektorial gambar dan jaringan saraf berulang untuk memecahkan kode representasi tersebut ke dalam kalimat bahasa alami -Mampu memvalidasi penggunaan perhatian dengan kinerja mutakhir pada tiga kumpulan data benchmark: Flickr8k, Flickr30k dan MS COCO.	Pada proses peningkatan kualitas pembuatan teks dengan jaringan saraf berulang untuk memecahkan kode representasi dalam bahasa alami.

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
3	Justin Johnson, Andrej Karpathy, Li Fei-Fei, 2016	*Fully Convolutional Localization Networks for Dense Captioning*	-Metode *dense captioning task* dengan sistem computer vision untuk mendeskripsikan wilayah yang menonjol dalam gambar dalam bahasa alami. pemberian teks padat -Generalisasi deteksi objek ketika deskripsi terdiri dari satu kata, dan Pembuatan Teks Gambar ketika satu wilayah yang diprediksi mencakup gambar penuh.	-Menggunakan arsitektur *Fully Convolutional Localization Network (*FCLN), dan Jaringan Konvolusional dengan lapisan lokalisasi padat baru, dan model bahasa Jaringan Neural Berulang yang menghasilkan urutan label. -Jaringan di evaluasi pada kumpulan data Genom Visual, yang terdiri dari 94.000 gambar dan 4.100.000 teks yang didasarkan pada wilayah.	Relevansinya dalam hal mendeskripsikan objek yang dominan dalam sebuah wilayah visual, yang selanjutnya dideskripsikan dalam teks yang padat (*dense captioning task*)
4	Guo, Yanming Liu, Yu Oerlemans, Ard Lao, Songyang Wu, Song Lew, Michael S, 2016	*Deep Learning for Visual Understanding:A Review*	-Skema *Deep Learning* untuk berbagai aplikasi *computer vision*, yaitu klasifikasi gambar, deteksi objek, pengambilan gambar, segmentasi semantik, dan estimasi pose manusia.	*-Deep learning algorithms* dibagi dalam 4 katagori: *Convolutional Neural Networks, Restricted Boltzmann Machines, Autoencoder* dan *Sparse Coding* - Sebuah klasifikasi gambar dari AlexNet, dimana setiap gambar memiliki 1 label kebnenaran dengan teknik probabilitas, dalam algoritma yang berbasis CNN kedalam 5 area yaitu: *image classification,object detection,image retrieval, semantic segmentation, human pose estimation.*	Relevansinya yaitu pembelajaran mendalam (*deep learning*) yang diadopsi secara luas dalam *computer vision*, seperti klasifikasi gambar, deteksi objek, pengambilan gambar dan segmentasi semantik, dan estimasi pose manusia,
5	Philip Kinghon, L.Zhang,L.S	*A Region-Based Image Caption Generator with*	Pembuatan Keterangan Gambar pada wilayah visual berbasis *deep learning*	-Arsitektur Convolutional Neural Network (CNN) dengan Recurrent Neural Network (RNN). berfungsi dengan baik dalam	Relevansi dalam implementasi deep learning dalam mengekstrak fitur gambar (CNN) dan membentuk pola

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
	hao, 2016	*Refined Description*	dengan deskripsi yang disempurnakan	menghasilkan teks pendek dan tingkat tinggi. Dalam metodologi ini CNN mengekstrak fitur gambar dari keseluruhan gambar, dengan membentuk vektor fitur yang sangat besar, yang terdiri dari antara dimensi 128 dan 4096, yang berarti metode *Machine Learned* yang khas tidak dapat mengatasi skala terutama dalam skenario waktu nyata atau dunia nyata. Metode ini dipasangkan dengan RNN yang mempelajari urutan dan pola dalam teks gambar yang menghasilkan teks kalimat.	teks gambar yang menghasilkan teks kalimat (RNN)
6	Peratham Wiriyatham mabhum, Douglas, Summer-Stay, Cornelia Fermuller, Yiannis Aloimonos, 2016	Computer Vision and Natural Language Processing Apprioaches in Multimedia and Robotics	Mengintegrasikan visi komputer dan pemrosesan bahasa alami, yang meliputi atribut visual, teks gambar, teks video, sebagai tema terpadu dari semantik distribusi. Analog semantik distribusi dalam visi komputer dan pemrosesan bahasa alami masing-masing sebagai penyematan gambar dan penyematan kata.	Konsep hubungan antara *Natural Language Processing* (NLP) dengan *computer vision* dalam aplikasi multimedia dan robotics. Pemrosesan Bahasa Alami (NLP) dapat diringkas menjadi konsep mulai dari sintaks, semantik dan pragmatik di tingkat atas untuk mencapai komunikasi. -Sintaks mencakup morfologi (studi tentang bentuk kata) dan komposisionalitas (komposisi unit bahasa yang lebih kecil seperti kata hingga unit yang lebih besar seperti frasa atau kalimat). -Semantik adalah studi tentang makna, termasuk menemukan hubungan antara kata, frasa, kalimat atau wacana. -Pragmatik mempelajari bagaimana makna	-Relevansinya bagaimana mengintegrasikan visi dan bahasa. oleh suatu sistem, pada level pengetahuan, sistem harus mampu menjawab pertanyaan terkait kemampuan untuk mengenali objek dan alat, manusia dan bagiannya, serta tindakan khusus dan peristiwa. -Model Semantik Distribusi (DSMs) menggunakan ruang vektor dan propertinya untuk memodelkan makna. Sebuah ruang vektor semantik merepresentasikan sebuah kata sebagai titik data dan mengkodekan kesamaan dan keterkaitan antara kata-kata dalam hal pengukuran antara

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
				berubah dengan adanya konteks tertentu.	titik-titik data tersebut.
7	Doersch, Carl Zisserman, Andrew, 2017	*Multi-task Self-Supervised Visual Learning*	Evaluasi metode pada klasifikasi image dengan menggunakan arsitektur ResNet-101 dengan Teknik Multiple selfsupervised tasks pada pelabelan secara manual.	Metode dengan Teknik multitask untuk menggabungkan beberapa Langkah labelling pada representasi visual secara manual yaitu: klasifikasi ImageNet, PASCAL VOC dan prediksi kedalaman NYU.	Pada penelitian menggunakan metode multi task deep learning, referensi memberikan masukan Teknik multiple tasking dalam proses representasi visual berupa deteksi objek dan *image recognition*
8	Chen, Bor Chun Ghosh, Pallabi Morariu, Vlad I.Davis, Larry S, 2017	*Detection of Metadata Tampering Through Discrepancy between Image Content and Metadata Using Multi-task Deep Learning*	*Convolutional Neural Network* (CNN) dengan *joint multi task learning*, untuk memberikan informasi meterorologi dan prediksi dari semua kondisi cuaca berdasarkan pengamatan visual	Teknik *CNN Regretion* dengan model AlexNet berbasis regresi L2, dengan multi task learning pada objek gambar yang menghasilkan metadata *tempering detection* untuk memprediksi informasi meteorologi pada klasifikasi Temperatur, Kelembaban dan Kondisi Cuaca dengan metode *Multi Task learning.*	Relevansinya yaitu pada model gabungan /multi task dari berbagai jenis metadata gambar, hal ini terkait dengan langkah menentukan area objek, grafik adegan dan *region caption* pada proses context citra gambar dari frame video.
9	Ruder, Sebastian, 2017	*An Overview of Multi-Task Learning in Deep Neural Networks*	Gambaran *Neural Network* pada *multi task learning*,dalam pemrosesan bahasa alami dan pengenalan ucapan	-Teknik Fast R-CNN (*Region-based Convolutional Neural Network*) yang dibangun untuk deteksi dan pengklasifikasian objek secara efesien dalam jaringan konvolusional. -Fast R-CNN dapat melatih jaringan VGG16 tiga kali lebih cepat dan menguji 10 kali lebih cepat -Fast R-CNN diimplementasikan dengan	Relavansinya dalam proses pengklasifikasian objek dan teknik pengujian dengan tingkat akurasi yg lebih tinggi. Lainnya yaitu sebagai referensi implementasi Python dan C++ dalam open source github.com

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
				Pyhton dan C++ dibawah lisensi MIT open-source http://github.com	
10	Cipolla, Roberto Gal, Yarin Kendall, Alex, 2018	*Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics*	Pembelajaran *multi task learnin*g yang mempertimbangkan beberapa kerugian karena dilaksanakan dengan besaran sekala yang berbeda secara bersamaan. Dengan mode; rergresi pada kedalaman pixel, semantic dan segmentasi instance dari gambar input.	Gambaran Multi task learning dengan *Homoscedastic Uncertainty* yang dibagi menjadi 2: 1. Ketidakpastian epistemik adalah ketidakpastian dalam model, yang menangkap apa yang tidak diketahui oleh model objek karena kurangnya data pelatihan. Ini dapat dijelaskan dengan peningkatan data pelatihan. 2. Ketidakpastian Aleatoric menangkap ketidakpastian kami sehubungan dengan kemampuan untuk mengamati semua variabel penjelas dengan presisi yang meningkat.	Relevansinya dalam pengaturan *multi task*, menunjukkan bahwa terdapat ketidakpastian tugas dari masing-masing komponen dalam hubungan relatif antar tugas, yang mencerminkan ketidakpastian yang melekat pada tugas regresi atau klasifikasi. Ini juga akan tergantung pada representasi tugas atau unit ukuran. Dimana pada referensi ini menggunakan model ketidakpastian homoscedastic sebagai dasar untuk pembobotan kerugian dalam masalah pembelajaran *multi task*.
11	Pugaliya, Hemant Saxena, Karan Garg, Shefali Shalini, Sheetal Gupta, Prashant Nyberg, Eric	*Multi-task learning for filtering and re-ranking answers using language inference and question entailment*	Arsitektur *deep learning* paralel seperti BERT dan MT-DNN yang merupakan metode pembelajaran untuk kumpulan data besar yang mampu bekerja dengan baik pada banyak tugas dan kumpulan data klasifikasi dokumen yang besar.	Mekanisme pengaturan Dataset untuk pemeringkatan ulang dan pemfilteran dengan MediQA. Dataset pelatihan terdiri dari 208 objek sedangkan penilaian dan set data uji memiliki masing-masing data uji 25 dan data latih 150. Setiap objek memiliki hingga 10 kandidat dataset	Relevansinya pada penyempurnaan klasifikasi dokumen besar dalam bentuk dataset dengan fitur multi task yaitu *pre trained* model dan *re rangking*

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
	Mitamura, Teruko, 2019				
12	Hessel, Matteo Soyer, Hubert Espeholt, Lasse Czarnecki, Wojciech Schmitt, Simon van Hasselt, Hado, 2019	*Multi-task deep reinforcement learning with PopArt*	Konsep multi tugas *multitask learning* dengan algoritma pembelajaran yang dapat menyelesaikan tidak hanya satu tetapi beberapa tugas berurutan sekaligus dengan permasalahan keseimbangan antara tugas yang diselesaikan dengan sumber daya terbatas	Penggunaan algoritma kritikus skala-invarian yang memungkinkan peningkatan kinerja secara signifikan dalam pengaturan pembelajaran *multi task deep learning* dengan mengevaluasi pendekatan dalam dua tolak ukur multi-tugas yaitu Atari-57 dan DmLab-30, masing-masing berdasarkan Atari dan DeepMind Lab.	Referensi untuk menevaluasi kinerja pada tolak ukur dalam menentukan skor agregat pada banyak kasus *multitask*, dengan menormalkan skor Eksperimen pada penelitian ini menggunakan data latih berbasis populasi (PBT) untuk mengadaptasi *hyperparameter* saat proses berlangsung
13	Ciaparrone, Gioele Luque Sánchez, Francisco Tabik, Siham Troiano, Luigi Tagliaferri, Roberto	*Deep learning in video multi-object tracking: A survey*	-Dalam penelitian ini dianalisa metode multi object *tracking* berbasis *deep learning*. -Menyelidiki fungsionalitas jaringan dengan mengklasifikasikan metode menjadi tiga kategori yaitu deskripsi menggunakan fitur dalam, penyematan jaringan dalam, dan konstruksi jaringan dalam ujung ke	-Analisis pelacakan multi-objek berbasis *deep learning* dengan Arsitektur Siamese CNN dibagi menjadi 3 metode: 1. Peningkatan pelacakan multi-objek menggunakan fitur jaringan dalam (*deep feature*), di mana fitur semantik diekstraksi dari jaringan saraf tiruan 2. Pelacakan multi-objek dengan penyematan jaringan dalam (*network embedding*) 3. Pelacakan multi-objek dengan	Releansinya yaitu dalam metode pelacakan multi objek berbasis *deep learning*. Selain CNN digunakan kombinasi RNN dan LSTM yang memilikim kinerja lebih baik dalam mengintegrasikan fitur tampilan dan probabilitas pencocokan antara setiap objek yang dideteksi, yang di evaluasi menggunakan *mean square error*.

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
	Herrera, Francisco, 2020		ujung. Kemudian struktur jaringan yang dalam dalam metode ini, dipergunakan untuk masalah pelacakan multi-objek.	pembelajaran jaringan dalam dari ujung ke ujung (*end to end deep network learning)*	
14	Liu, Xiaodong He, Pengcheng Chen, Weizhu Gao, Jianfeng, 2020	*Multi-task deep neural networks for natural language understanding*	*Multi-Task Deep Neural Network* (MT-DNN) untuk representasi pembelajaran dalam pemahaman bahasa alami / multiple *Natural Language Understanding* (NLU). MT-DNN tidak hanya memanfaatkan data lintas tugas dalam jumlah besar, tetapi juga memanfaatkan efek regularisasi yang mengarah ke representasi yang lebih umum untuk membantu beradaptasi dengan tugas dan domain baru.	-Model *Multi-Task Deep Neural Network* MT-DNN untuk menggabungkan pembelajaran multi-tugas dan model bahasa pra-pelatihan untuk pembelajaran representasi bahasa dengan implementasi PyTorch dari BERT5 -MT-DNN memperoleh hasil mutakhir baru pada sepuluh tugas NLU di tiga tolak ukur popular yaitu SNLI, SciTail, dan GLUE. MT-DNN juga menunjukkan kemampuan generalisasi yang luar biasa dalam eksperimen adaptasi domain dengan cara menggabungkan struktur linguistik teks dengan cara yang lebih eksplisit.	Relevansi dalam penelitian ini yaitu dalam proses menggabungkan *multi task learning* dari model bahasa kedalam bentuk representasi bahasa.
15	Cipolla, Roberto Gal, Yarin Kendall, Alex, 2018	*Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics*	Metode *multi task deep learning* dengan arsiterktur gambar RGB monokuler tunggal sebagai input dan menghasilkan klasifikasi berdasarkan piksel,	Pada jaringan *multi task* terdapat decoder yang berfungsi untuk mempelajari pemetaan dari fitur-fitur bersama ke sebuah keluaran. Setiap dekoder terdiri dari lapisan konvolusional 3 × 3 dengan ukuran fitur keluaran 256, diikuti oleh lapisan 1 × 1	Dalam penentuan pembobotan multi task yang optimal menggunakan ide-ide dari pemodelan probabilistik

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
			segmentasi semantik instance, dan perkiraan kedalaman per piksel.	yang menurunkan keluaran tugas. Segmentasi Semantik menggunakan kerugian cross-entropy untuk mempelajari probabilitas kelas berdasarkan piksel, meratakan kerugian pada piksel dengan label semantik di setiap tumpukan mini.	
16	Hessel, Matteo Soyer, Hubert Espeholt, Lasse Czarnecki, Wojciech Schmitt, Simon van Hasselt, Hado, 2019	*Multi-task deep reinforcement learning with PopArt Deep learning in video multi-object tracking: A survey*	*Multi-task deep reinforcement learning* dengan PopArt Deep learning dalam pelacakan video multi-objek	Normalisasi adaptif PopArt dapat dikombinasikan dengan penelitian lain dalam RL multi-tugas, yang sebelumnya tidak berskala efektif untuk sejumlah besar tugas yang berbeda Kombinasi PopArt-IMPALA dengan bentuk-bentuk pengambilan sampel aktif karena memungkinkan penggunaan yang lebih efisien dari pembuatan data paralel, dengan memfokuskannya pada tugas yang paling sesuai untuk pembelajaran.	Relevansi dalam pengaturan parallel learning untuk mengurangi interferensi pada proses multi task. Algoritma dibuat untuk memiliki kinerja yngh kompleks pada satu waktu.
17	Tjatur Kandaga Gautama, Antonius Hendrik, Riskadewi, 2016	Pengenalan Objek pada Computer Vision dengan Pencocokan Fitur Menggunakan Algoritma SIFT Studi Kasus: Deteksi Penyakit Kulit Sederhana	Deteksi citra dengan algoritma Scale invariant Feature Transform (SIFT), untuk menentukan jenis penyakit kulit yang dicocokkan ndengan dataset citra deteksi tepi.	Melakukan manipulasi (pre-processing) yang dilakukan pada citra digital a. Pengurangan noise (noise reduction) dengan menggunakan filter Gaussian b. Pemecahan (split) citra warna ke dalam masing-masing color channel (merah, hijau, dan biru) 8 bit 3. Basis data citra untuk penyakit kulit berhasil dibangun. Basis data ini berisi	Referensi untuk proses *Image acquisition*, prosedur manipulasi (*pre-processing*) dengan *noise reduction* dan pemisahan (*split*) citra warna ke masing-masing color c*hannel*

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
				pengetahuan tentang penyakit kulit dan citra hasil deteksi tepi. -Penelitian ini menghasilkan perangkat lunak sistem pakar diagnosis penyakit kulit. Perangkat lunak menerima input berupa citra digital hasil deteksi tepi dari penyakit kulit dan informasi mengenai penderita penyakit kulit seperti jenis kelamin, usia, dan warna penyakit kulit	
18	Somak Aditya, Yezhou Yang, Chitta Baral, 2017	Image Understanding using Vision and Reasoning Through Scene Description Graph	Dua tugas mendasar dalam memahami gambar menggunakan teks adalah pembuatan teks dan menjawab pertanyaan visual. -Struktur pengetahuan yang digunakan yaitu *Scene Description Graph* (SDG), karena ini adalah grafik berlabel terarah, mewakili objek, tindakan, wilayah, serta atributnya, bersama dengan konsep dan semantik yang disimpulkannya.	Membangun SDG (*Scene Description Graph)* dari deteksi: -Phase Preprocessing -Knowledge extraction and storage dengan membangun knowledge base dan Baysian Network -Inferensi melalui pengetahuan dan penalaran -Pembuatan teks adalah tugas menghasilkan kalimat deskriptif yang relevan dari gambar; relevansi dan ketelitian menjadi dua kriteria berbeda, yang dengannya kualitas teks dapat dinilai. Dengan menggunakan image understanding untuk menguji relevansi dan ketelitian yang dihasilkan	Relevansi pada proses SDG (*Scene Description Graph)* sehingga mampu membangun knowledge untuk menghasilkan kalimat deskriptif dari sebuah gambar, dengan mengadopsi 2 eksperimen untuk mengevaluasi secara kualitatif dan evaluasi penyelarasan image-kalimat.

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
19	Dewi, Syarifah Rosita, 2018	Deep Learning Object Detection Pada Video Menggunakan Tensorflow Dan Convolutional Neural Network	Metode *Deep learning* yang digunakan untuk pengenalan dan klasifikasi objek adalah *Convolutional Neural Network* karena banyak digunakan pada penelitian terdahulu dan menghasilkan hasil yang signifikan dalam pengenalan citra. Pada penelitian ini dilakukan pengenalan objek meja dan kursi motif ukiran Jepara menggunakan framework Tensorflow dengan dataset sebanyak 500 gambar.	- Model yang terbentuk adalah model hasil training dengan jumlah 250000 steps dengan 2 batch size yaitu berupa graph inference yang terdiri dari file checkpoint, frozen_inference_graph.pb, dan terdapat 3 file model-ckpt yang masing-masing berekstensi .data-00000-of-00001, .index, dan .meta yang diletakkan dalam satu direktori yang sama dan digunakan untuk melakukan pengujian model pada saat melakukan pendeteksian objek. - Tingkat akurasi model yang didapatkan dari hasil pendeteksian klasifikasi citra meja dan kursi motif ukiran Jepara pada suatu citra digital menggunakan Convolutional Neural Network berkisar antara 70% hingga 99%	Referensi pada penelitian ini yaitu pengunaan algoritma CNN untuk mendeteksi dan mengklasifikasi objek dengan jumlah 502 dataset, 80% untuk training dan 20% untuk testing. Penelitian ini menghasilkan akurasi sekitar 99% dengan batch 4 dan dengan 100.000 steps sampai proses training berhasil mendeteksi sebuah objek
20	Donghyeop Shin, Incheol Kim, 2018	Deep Image Understanding Using Multilayer Context	- Grafik pemandangan dan keterangan bahasa alami memiliki karakteristik umum yang dihasilkan dengan mempertimbangkan objek dalam gambar dan hubungan antar objek. - penelitian ini mengusulkan model *deep neural network* bernama *Context-based Captioning and Scene Graph Generation Network*	-Metode memecahkan masalah pemahaman citra tingkat tinggi dengan menggunakan model pemahaman citra tingkat rendah yang ada. Metode ini dapat digunakan untuk menyelesaikan masalah yang menuntut pemahaman citra tingkat tinggi seperti pemahaman ekspresi referensi, pengambilan citra, dan menjawab pertanyaan visual. Model C2SGNET (*Context-based Captioning and Scene Graph Generation Network)*, yang dapat secara bersamaan	Relevansinya yaitu pada proses hasil prediksi dan pelatihan model yang disusun menjadi tiga lapisan: deteksi objek, deteksi hubungan, dan lapisan pembuatan teks. Hasilnya diprediksi melalui empat langkah: proposal wilayah kandidat, ekstraksi fitur wilayah, ekstraksi fitur konteks, dan pembuatan grafik adegan dan pembuatan judul.

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
			(C2SGNet), yang secara bersamaan menghasilkan grafik adegan dan teks bahasa alami dari gambar.	menghasilkan grafik pemandangan dan keterangan bahasa alami dari gambar masukan untuk pemahaman gambar tingkat tinggi. Model ini menggunakan fitur-fitur yang terkait dengan setiap tugas sebagai informasi konteks, berdasarkan karakteristik bahwa grafik pemandangan dan keterangan bahasa alami dapat dihasilkan dari objek dan hubungan antar objek.	
21	Shuang Bai, Shan An, 2018	A Survey an Automatic Image Caption Generation	Metode jaringan saraf tiruan untuk pemberian keterangan gambar secara otomatis. Untuk mencapai tujuan pembuatan teks gambar, informasi semantik gambar perlu ditangkap dan diekspresikan dalam bahasa alami.	-Dengan kerangka encoder-decoder untuk menambahkan mekanisme perhatian ke model, sehingga model encoder-decoder mampu secara dinamis menampilkan daerah gambar yang menonjol selama proses pembuatan deskripsi gambar -Skor BLEU-n dan METEOR digunakan pada ketiga kumpulan data benchmark. -Model perhatian semantik untuk memasukkan isyarat visual kognitif ke dalam decoder sebagai panduan perhatian untuk captioning gambar. -Metode dievaluasi pada dataset Flickr30k dan MSCOCO, dengan skor	Relevansi pada hasil penelitian menunjukkan bahwa modifikasi yang sesuai dengan kerangka encoder-decoder dasar dengan memperkenalkan mekanisme perhatian dapat meningkatkan kinerja captioning gambar secara efektif.

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
				BLEU-n dan METEOR	
22	Ghosh, Swarnendu Das Nibaran, Das Ishita, Maulik, Ujjwal, 2019	Understand Deep Learning Technique For Image Segmentation	Pendekatan segmentasi citra tradisional, yang menjelaskan tentang pengaruh pembelajaran mendalam terhadap domain segmentasi citra. Algoritme segmentasi utama telah dikategorikan secara logis dengan paragraf yang didedikasikan untuk kontribusi uniknya.	-Metode CNN dasn RNN melalui proses *pooling layer*, *Fully Connected Layer* (FC), *activation function*, *batch normalization*, dan *dropout* (proses mengurangi koneksi yang berlebihan antara neuron dalam jaringan dengan mencegah koordinasi kompleks dalam data pelatihan), LaexNet, ZFNet, Google Net, VGG Net, ResNet - Jaringan saraf konvolusi (CNN), dianggap sebagai kelas jaringan saraf dalam, memiliki kemampuan tinggi untuk mendeteksi pola gambar yang banyak digunakan dalam algoritme *computer vision.*	Pendekatan yang secara langsung menggunakan jaringan neural dalam (terutama CNN) untuk mengklasifikasikan gambar dengan sub-fitur. Pendekatan menggunakan jaringan neural dalam lebih sebagai ekstraktor fitur untuk menemukan dan menyelaraskan bagian yang berbeda dari objek Pendekatan yang menggunakan beberapa jaringan neural dalam untuk membedakan dengan lebih baik antara gambar visual yang sangat kecil.
23	Raimonda Staniute, Dmitrij Sesok, 2019	A Systematic Literature Review on Image Captioning	Metode Systematic Literature Review (SLR) dan NLP (*Natural Language Processing*) yang komprehensif memberikan gambaran singkat tentang perbaikan teks gambar. Fokus utama penelitian ini adalah untuk menjelaskan teknik yang paling umum dan tantangan terbesar dalam	Proses menggunakan komponen sebagai berikut: -Feature extractors: AlexNet, VGG-16 Net, ResNet,GoogleNet (including all nine Inception models); DenseNet; -Language models: LSTM, RNN, CNN, cGRU, TPGN; -Methods: Encoder-decoder, attention mechanism, Novel objects, semantics	Referensi tentang image captioning dengan dataset MS COCO dan Flick 30K untuk evaluasi model

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
			pembuatan teks gambar	-Results on datasets: MS COCO, Flickr30k	
24	Jácome-Galarza, Luis Roberto Realpe-Robalino, Miguel Andrés Chamba-Eras, Luis Antonio Viñán-Ludeña, Marlon Santiago, 2020	Computer Vision for Image Understanding A Comperensive Review	Kombinasi *Natural Language Processing* (NLP) menggunakan *Recurrent Neural Net- works dan Long Short-Term Memory*) ditambah *Image Understanding* (menggunakan *Convolutional Neural Network*s) dapat menghadirkan jenis aplikasi baru tentang konten gambar dan video.	Pada penelitian ini menggunakan framework sebagai berikut: - *Image Segmentation* - *Event recognition.* - *Scene recognition* - *Fine-grained recognition* - *Action Classification.* - *Action Localization* - *Object category recognition.* - *BoW (Bag of Words).* - *Hierarchical model.* - *ConvNets.* Model dan dataset untuk deskripsi gambar yang tersedia untuk publik di repositori menggunakan github	Relevansi yaitu dalam Pelabelan adegan dalam kaitannya dengan computer vision yang membutuhkan penggunaan lokal fitur diskriminatif dan informasi konteks global. Proses ini juga mengadopsi jaringan saraf konvolusional berulang (RCNN)
25	Haneol Jang, Jong -Uk Hou,2020	Exposing Digital Image Forgeries by Detecting Contextual Abnormality Using Convolutional Neural Networks	Merode menggunakan jaringan saraf dalam untuk mendeteksi manipulasi gambar berdasarkan kelainan kontekstual. - Metode yang diusulkan mendeteksi kelas dan lokasi objek menggunakan detektor objek terkenal seperti jaringan	-Model (CL-CNN) yang dapat memberikan informasi kontekstual sebelumnya langsung mempelajari kombinasi label gambar. -Model terlatih memberikan prior kontekstual berdasarkan CNN. Metode yang diusulkan pertama kali menggunakan detektor objek terkenal seperti R-CNN [30] untuk mendeteksi kelas dan lokasi objek	Relevansinya yaitu detektor objek berbasis wilayah yang digunakan dalam penelitian ini, dimana konteks antara objek dan latar belakangnya dievaluasi secara langsung. Oleh karena itu, untuk meningkatkan akurasi studi dengan menggabungkan pembelajaran mendalam berdasarkan klasifikasi adegan dalam sebuah

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
			saraf konvolusional berbasis wilayah (R-CNN) dan mengevaluasi skor kontekstual sesuai dengan kombinasi objek, konteks spasial objek, dan posisi benda.	dan kemudian mengevaluasi skor kontekstual berdasarkan kombinasi objek. - **E**ksperimen ini dilakukan dengan menggunakan gambar sampel yang dikumpulkan dari Microsoft COCO: Common Objek dalam Konteks Dan menggunakan 65.268 gambar multi-kategori, yang dibagi menjadi 80 kategori, untuk melatih CL-CNN. - Sebuah set positif dibangun menggunakan informasi label dari gambar multi-kategori	*region*/wilayah
26	Z.Aini, F.Hutapea, N.Ramadhanie, 2020	Implementasi Sistem Pegawas CCTV Di Kota Tanjungpinang (Studi Kasus Dinas Perhubungan)	Masalah yang patut diperhatikan di kota besar saat ini adalah masa lalu lintas. Hal tersebut bisa dilihat dari meningkatnya angka pelanggaran dan kecelakaan lalu lintas lintas setiap tahunnya. Salah satu upaya untuk mengatasi permasalahan pelanggaran lalulintas yaitu denhgan perangkat CCTV	Manfaat yang dirasakan dalam sistem pengawasan CCTV lalu lintas di Kota Tanjungpinang belum optimal karena kenyataannya hanya dapat mempercepat penertiban lalu lintas, dan mengurangi kemacetan lalu lintas. disamping itu target yang hendak dicapai yaitu masyarakat lebih berbudaya dalam berlalu lintas karena selama ini budaya masyarakat dalam berlalu lintas masih kurang.	Dalam menentukan objek penelitiannya, yaitu fenomena banyaknya pelanggaran yang dilakukan masyarakat, sehingga pihak berwajib bisa menentukan jenis pelanggaran melalui lamera CCTV dengan deskripsi kontekstual secara otomatis.
27	D.Ariyoga, R.Rahmadi,	Penelitian Terkini Tentang Sistem	Mertode penelitian menggunakan algoritma-	-Metode-metode yang digunakan memperoleh nilai akurasi dan ketepatan	Relevansinya yaitu dari kajian literatur yang telah dilakukan, dapat

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
	R.Rajagede, 2021	Pendeteksi Pelanggaran Lalu Lintas Berbasis Deep Learning: Sebuah Kajian	algoritma Deep Learning seperti YOLO, CNN, Faster R-CNN dan metode- metode dari Image Processing seperti Haar-like Feature dan Edge Detection serta metode Evolutionary Programming seperti Genetic Algorithm dalam pengembangan masing-masing sistemnya.	yang cukup dan bahkan sangat baik. Namun, terdapat beberapa kelemahan pada pendeteksian kendaraan dan pelanggaran pada keadaan minim cahaya. Selain itu, ditemukan juga bahwa banyaknya data set dan beragamnya sudut pandang pada gambar pada proses pelatihan berpengaruh pada kecepatan dan nilai akurasi hasil model. -Algoritma dan metode deteksi yang digunakan pada sebelas literatur yang dikaji, yaitu YOLO, CNN, Faster R-CNN, Haar-like Feature, Edge Detection, dan Genetic Algorithm. Rata-rata akurasi yang dicapai dari penggunaan berbagai metode pada penelitian-penelitian diatas juga sudah cukup baik. Di antara metode lainnya, CNN dan YOLOv3 menjadi metode dan algoritma yang paling umum dipakai dengan hasil akurasi yang paling tinggi, yaitu di atas 90%	disimpulkan bahwa perkembangan teknologi *Object Detection, Computer Vision, Image Processing, dan Deep Learning* sangat bisa dimanfaatkan untuk membuat sistem untuk deteksi kendaraan pada lalu lintas dan deteksi otomatis pelanggaran pada lalu lintas.
28	Z.Zang, 2010	*Research on the taxi traffic accident and violation identification model*	- Sistem pengawasan lalu lintas harus efektif dengan identifikasi yang akurat. - Dirancang suatu model matematika identifikasi pelanggaran yang akurat. dengan menggunakan algoritma pendeteksian secara	*Object recognition* dengan Hopfield's Neural Model, sistem pengenalan pola hybrid berdasarkan *Genetic Algorithm* (GA) dengan model saraf *Hopfield (HP*) yang bahkan dapat mengenali pola yang dideformasi oleh transformasi yang disebabkan oleh rotasi, penskalaan, atau terjemahan, secara tunggal atau dalam	Relevansinya yaitu pada saat proses object recognition di jalan raya untuk menentukan objek pelanggaran lalu lintas.

Tabel 2.4 Matrix State of The Art

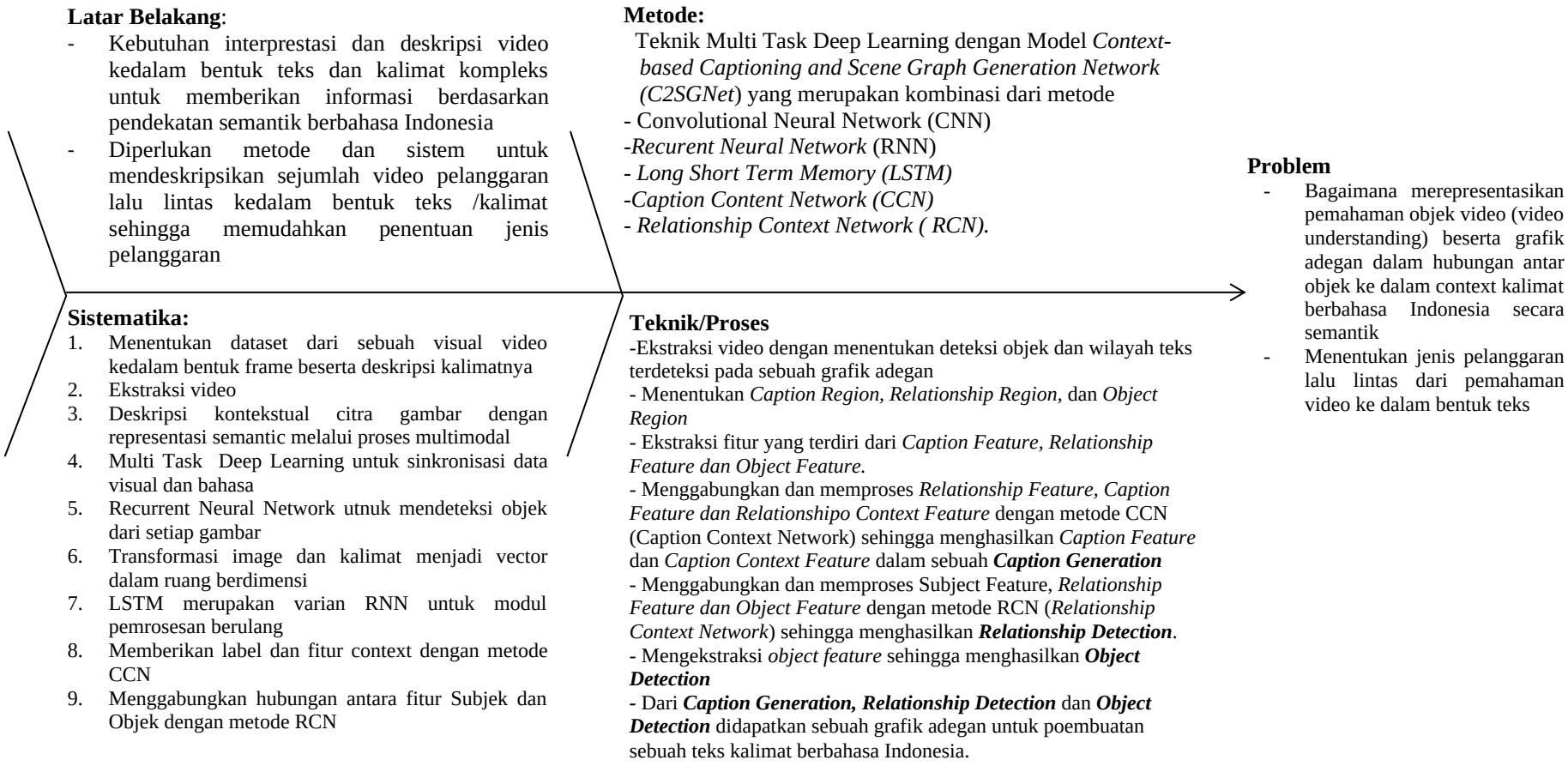
No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
			efektif mengawasi taksi, dengan demikian tingkat kecelakaan lalu lintas taksi dapat diturunkan.	kombinasi.	
29	K. Klubsuwan, W. Koodtalang and S. Mungsing, 2013	*Traffic Violation Detection Using Multiple Trajectories Evaluation of Vehicles*	-Metode baru untuk deteksi pelanggaran lampu merah menggunakan kendaraan yang bergerak di wilayah yang diinginkan dan dikombinasikan dengan evaluasi perilaku lintasan beberapa kendaraan menggunakan mean square displacement (MSD) untuk mendeteksi kedua pelanggaran tersebut.	Teknik pemrosesan gambar hanya untuk mendeteksi sinyal lalu lintas tanpa bantuan sistem lain. Sistem dapat mendeteksi lampu merah dan pelanggaran perubahan jalur dan manajemen lalu lintas, dengan menggunakan *Neural Network* Model.	Relevansinya dalam menentukan jenis pelanggaran lalu limntyas pada lokasi traffic light.
30	Singh, Vedant Unadkat, Vyom Kanani, Pratik, 2019	*Intelligent traffic management system*	-Menyimpan dan memproses data dalam jumlah besar secara efisien menggunakan teknik seperti Deep Learning dan Computer Vision	Sistem otomatis untuk mendeteksi Pelanggaran Lalu Lintas menggunakan YOLOv3 untuk mendeteksi dan melacak kendaraan serta menyimpan snapshot jika terjadi pelanggaran,yang disimpan dalam model ROI (*Region of Interest*)	Untuk menentukan jenis pelanggaran yang tertangkap melalui kamera CCTV
31	Ambekar, Shubham Dagade, Mangesh,	*Traffic signal violation monitoring through video*	- Deteksi pelanggaran lalu lintas model yang diusulkan menggunakan kamera web untuk mengidentifikasi nomor	- Metodologi ada untuk mengidentifikasi kendaraan menggunakan tag identifikasi frekuensi Radio (RF ID) - Pengenalan Nomor Melalui CNN -	Relevansi dalam proses identifikasi objek

Tabel 2.4 Matrix State of The Art

No.	Penulis dan Tahun	Judul Penelitian	Metode Penelitian	Fokus dan Hasil Penelitian	Relevansi Terhadap Penelitian yang Dilakukan
	2019	*surveillance using CNN)*	plat kendaraan yang dilanggar melalui jaringan saraf konvolusi, yang menggunakan teknik pencocokan pola untuk identifikasi nomor, sehingga untuk mendenda pelaku untuk hal yang sama	Sebagai sistem pengontrol pengenalan nomor menerima sinyal dari sistem pengawas penyeberangan zebra, itu mengambil bingkai untuk contoh itu dan kemudian mengubahnya menjadi gambar biner seperti yang disebutkan dalam algoritma 1.	

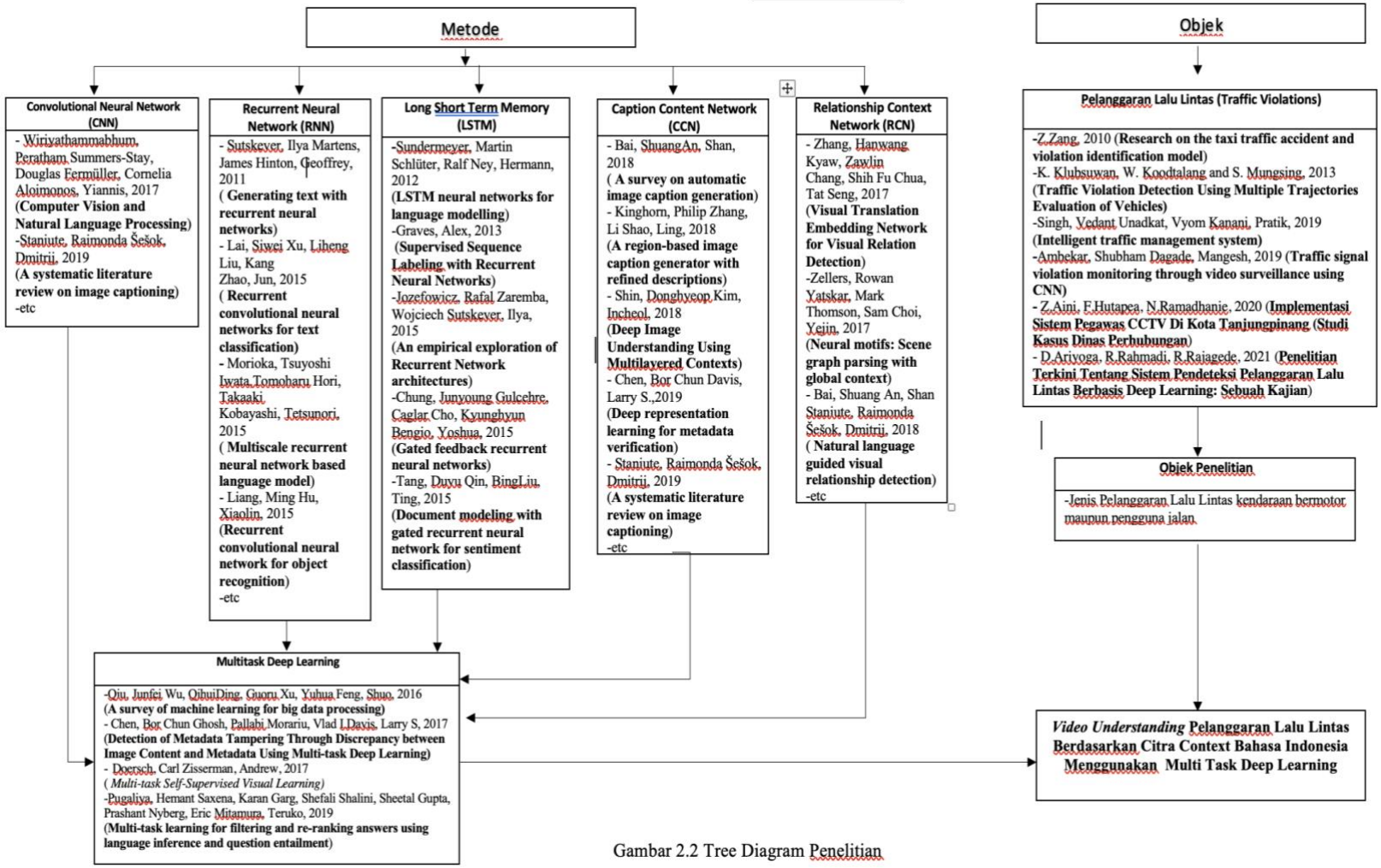
Pada gambar 2.1 berikut ini adalah fishbone dari penelitian dan pada gambar 2.2 adalah gambar Tree Diagram dari usulan penelitian ini.

FISHBONE PENELITAN



Gambar 2.1 Fishbone dari Penelitian

TREE DIAGRAM



Gambar 2.2 Tree Diagram Penelitian

2.2 Teori dan Metode

Grafika Komputer bertujuan menghasilkan citra (lebih tepat disebut grafik atau picture) dengan primitif-primitif geometri seperti garis, lingkaran dan sebagainya. Primitif-primitif geometri tersebut memerlukan data deskriptif untuk melukis elemen-elemen gambar. Contoh data deskriptif adalah koordinat titik, panjang garis, jari-jari lingkaran, tebal garis, warna, dan sebagainya. Grafika komputer memainkan peranan penting dalam visualisasi dan virtual reality.

Pengolahan Citra (image processing). Pengolahan Citra bertujuan memperbaiki kualitas citra agar mudah diinterpretasi oleh manusia atau mesin (dalam hal ini komputer). Teknik-teknik pengolahan citra mentransformasikan citra menjadi citra lain. Jadi, masukannya adalah citra dan keluarannya juga citra, namun citra keluaran mempunyai kualitas lebih baik daripada citra masukan. (Umam and Negara, 2016).

2.2.1 Computer Vision

Computer vision merupakan proses otomatis yang mengintegrasikan sejumlah besar proses untuk persepsi visual, seperti akuisisi citra, pengolahan citra, pengenalan dan membuat keputusan. *Computer vision* mencoba meniru cara kerja sistem visual manusia (*human vision*) yang sesungguhnya sangat kompleks. (Putra, 2010) Untuk itu, *computer vision* diharapkan memiliki kemampuan tingkat tinggi sebagaiman *human visual*. Kemampuan itu diantaranya adalah:

- Object detection, menentukan sebuah objek pada scene beserta batasannya
- Recognation, menempatkan label pada objek.
- Description, menugaskan properti kepada objek.

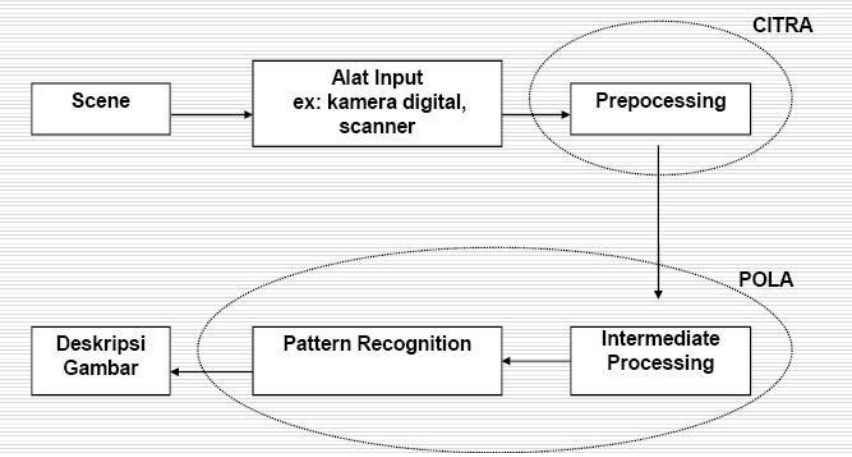
- 3D Inference, menafsirkan adegan 3D dari 2D yang dilihat.
- Interpreting motion, menafsirkan gerakan.

Computer vision menggabungkan kamera, komputasi berbasis *edge* atau *cloud*, perangkat lunak, dan kecerdasan buatan (AI) sehingga sistem dapat "melihat" dan mengidentifikasi objek. Intel memiliki portofolio teknologi penerapan AI yang lengkap, termasuk CPU untuk pemrosesan tujuan umum, *computer vision*, dan unit pemrosesan vision (VPU) untuk akselerasi. Sistem computer vision yang bermanfaat di berbagai lingkungan dapat mengenali objek dan orang dengan cepat, menganalisis demografi khalayak, memeriksa hasil produksi, juga banyak hal lainnya. *Computer vision* menggunakan pembelajaran mendalam untuk membentuk jaringan neural yang memandu sistem dalam pemrosesan dan analisis. Model *computer vision* yang telah sepenuhnya terlatih dapat mengenali objek, mendeteksi dan mengenali orang, bahkan melacak pergerakan. (Bregler, Covell and Slaney, 1997).

- **Proses dan Hirarki Pada Computer Vision**

  Ada terdapat 3 proses yang terjadi dalam computer vision, yaitu:

  - Memperoleh atau mengakuisisi citra digital.
  - Operasi pengolahan citra.
  - Menganalisis dan menginterpretasi citra dan menggunakan hasil pemrosesan untuk tujuan tertentu, misal memandu robot, mengontrol peralatan, dan sebagainya.



Gambar 2.3. Proses Pada Computer Vision

Hirarki pada *computer vision* ada 3 tahap, yaitu:

1) Pengolahan Tingkat Rendah (Image to image) → Menghilangkan noise, dan peningkatan gambar (*enchament image*).
2) Pengolahan Tingkat Menengah (Image to dimbolic) → Kumpulan garis / vektor yang merepresentasikan batas sebuah obyek PADA citra.
3) Pengolah Tingkat Tinggi (Simbolic to simbolic) → Representasi simbolik batas-batas obyek menghasilkan nama obyek tersebut.

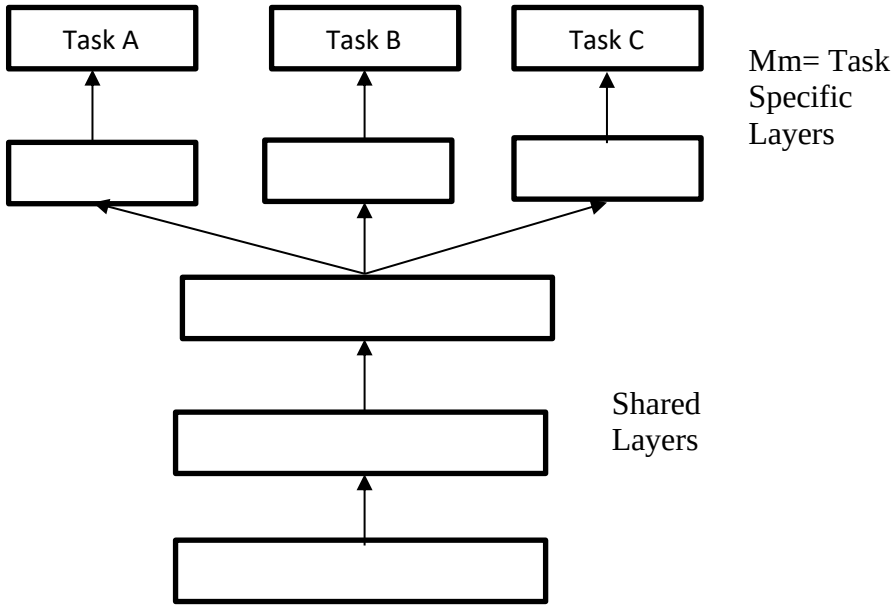
2.2.2 *Multi Task Learning*

Multi Task Learning (MTL) digunakan pada aplikasi pembelajaran mesin, dari pemrosesan bahasa alami, pengenalan ucapan, computer vision serta penemuan obat. MTL hadir dalam berbagai bentuk: pembelajaran bersama, belajar untuk belajar, dan belajar dengan tugas tambahan hanyalah beberapa nama yang telah digunakan untuk merujuk padanya. Umumnya, setelah mengoptimalkan lebih dari satu fungsi secara efektif melakukan pembelajaran multi-tugas (berbeda

dengan pembelajaran tugas Cara yang paling umum digunakan untuk melakukan pembelajaran multi-tugas di jaringan neural multi-tugas biasanya dilakukan dengan berbagi parameter lapisan tersembunyi baik keras atau lunak. Berbagi parameter keras adalah pendekatan yang paling umum digunakan untuk MTL di jaringan saraf. Ini umumnya diterapkan dengan berbagi lapisan tersembunyi di antara semua tugas, sambil mempertahankan beberapa lapisan keluaran khusus tugas.

- *Hard Parameter Sharing*

  Hard Parameter Sharing adalah pendekatan yang paling umum digunakan untuk MTL di jaringan saraf. Umumnya diterapkan dengan berbagi lapisan tersembunyi di antara semua tugas, dengan mempertahankan beberapa task – specific output layers.

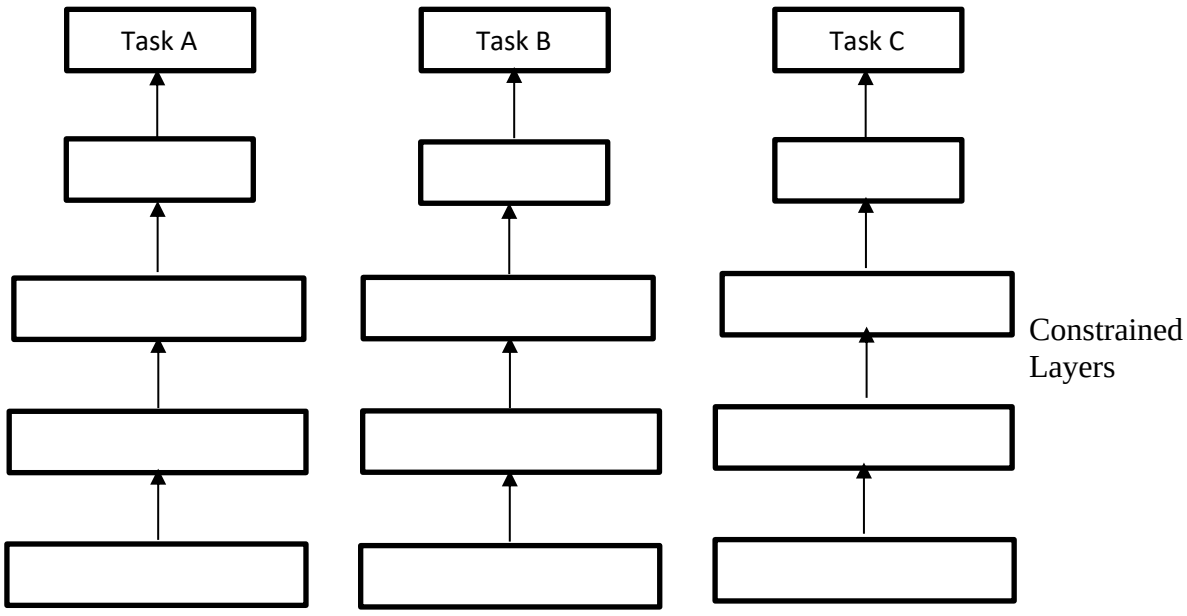


Gambar 2.4 *Hard parameter sharing for multi-task learning in deep neural networks*

Berbagi parameter keras sangat mengurangi risiko *overfitting*. Fakta menunjukkan bahwa risiko *overfitting* pada parameter bersama adalah urutan N - di mana N adalah jumlah tugas - lebih kecil daripada *overfitting* parameter khusus tugas, yaitu lapisan keluaran. Semakin banyak tugas yang kita pelajari secara bersamaan, semakin banyak model kita harus menemukan representasi yang menangkap semua tugas dan semakin sedikit peluang kita untuk melakukan *overfitting* pada tugas awal kita.

- *Soft Parameter Sharing*

  Di sisi lain, dalam berbagi parameter lunak, setiap tugas memiliki modelnya sendiri dengan parameternya sendiri. Jarak antar parameter model kemudian diatur untuk mendorong parameter menjadi serupa, misalnya menggunakan norma untuk regularisasi, dengan menggunakan norma jejak.

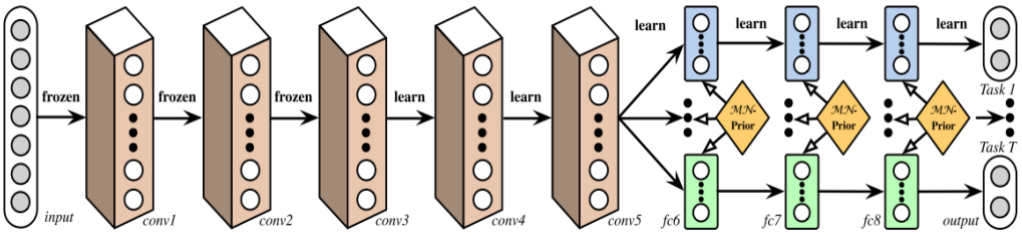


Gambar 2.5 *Soft parameter sharing for multi-task learning in deep neural networks*

Batasan yang digunakan untuk berbagi parameter lunak di jaringan neural dalam sangat terinspirasi oleh teknik regularisasi untuk MTL yang telah dikembangkan untuk model lain.

- *Deep Relationship Networks*

  Dalam MTL untuk computer vision, pendekatan sering kali berbagi lapisan konvolusional, sambil mempelajari lapisan terhubung penuh khusus tugas dengan memperbaiki model ini dengan *Deep Relationship Networks*. Selain struktur bersama dan lapisan khusus, yang dapat dilihat pada Gambar 2.6, mereka menempatkan matriks sebelumnya pada lapisan yang sepenuhnya terhubung, yang memungkinkan model untuk mempelajari hubungan antar tugas, mirip dengan beberapa model Bayesian.



Gambar 2.6
Deep Relationship Network dengan lapisan konvolusional bersama dan sepenuhnya terhubung dengan prior matriks (Ruder, 2017)

2.2.3 *Recurrent Neural Network (RNN)*

Pemodelan gambar generatif adalah masalah utama dalam pembelajaran tanpa pengawasan (*unsupervised learning*). Model kepadatan probabilistik dapat digunakan untuk berbagai macam tugas yang berkisar dari kompresi gambar dan bentuk rekonstruksi seperti gambar *inpainting* (misalnya, lihat Gambar 1) dan *deblurring*, hingga pembuatan gambar baru. Ketika model dikondisikan pada

informasi eksternal, aplikasi yang mungkin juga mencakup pembuatan gambar berdasarkan deskripsi teks atau simulasi bingkai masa depan dalam tugas perencanaan. Salah satu keuntungan besar dalam pemodelan generatif adalah bahwa secara praktis ada banyak sekali data gambar yang tersedia untuk dipelajari. Namun, karena gambar berdimensi tinggi dan sangat terstruktur, memperkirakan distribusi gambar alam sangatlah menantang. Salah satu kendala terpenting dalam pemodelan generatif adalah membangun model yang kompleks dan ekspresif. (Van Den Oord, Kalchbrenner and Kavukcuoglu, 2016).

Occluded completion original

Gambar 2.*7 Image completions sampled from a Pixel RNN.*

Salah satu kendala terpenting dalam pemodelan generatif adalah membangun model penyelesaian oklusi asli pada gambar 2.7. Penyelesaian gambar diambil sampelnya dari Pixel RNN. Pertukaran ini telah menghasilkan berbagai macam model generatif, masing-masing memiliki kelebihan. Sebagian besar pekerjaan berfokus pada model variabel laten stokastik seperti VAE (Rezende et al., 2014; Kingma & Welling, 2013) yang bertujuan untuk mengekstrak representasi yang bermakna, tetapi sering kali disertai dengan

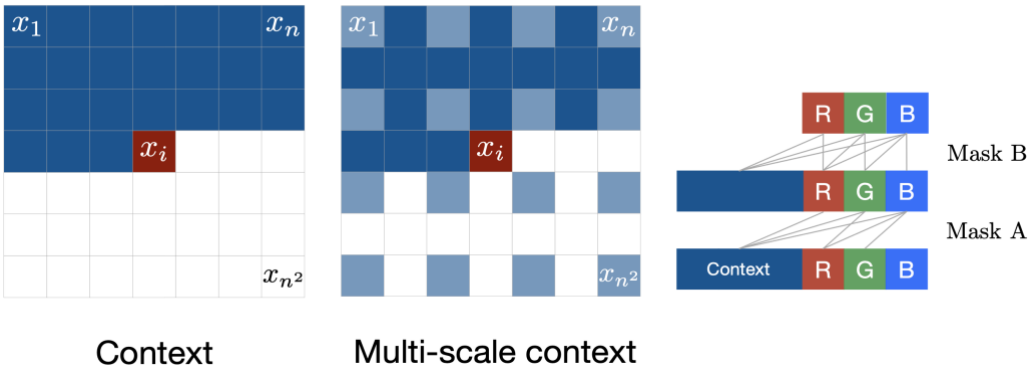
langkah inferensi yang sulit yang dapat menghambat kinerja mereka. Salah satu pendekatan efektif untuk memodelkan distribusi gabungan piksel dalam gambar dengan tepat adalah dengan menampilkannya sebagai produk distribusi bersyarat, pendekatan ini telah diadopsi dalam model autoregresif seperti NADE (Larochelle & Mur-ray, 2011) dan jaringan kepercayaan sigmoid yang terlihat sepenuhnya (Neal, 1992). Faktorisasi mengubah masalah pemodelan gabungan menjadi masalah urutan, di mana proses memprediksi piksel berikutnya berdasarkan semua piksel yang dihasilkan sebelumnya. Tetapi untuk memodelkan korelasi yang sangat nonlinier dan jarak jauh antara piksel dan distribusi bersyarat kompleks yang dihasilkan, diperlukan model urutan yang sangat ekspresif.

Recurrent Neural Networks (RNN) adalah model canggih yang menawarkan parametrisasi bersama yang ringkas dari rangkaian distribusi bersyarat. RNN telah terbukti unggul dalam masalah urutan keras mulai dari generasi tulisan tangan (Graves, 2013), hingga prediksi karakter (Sutskever et al., 2011) dan hingga terjemahan mesin (Kalchbrenner & Blunsom, 2013). RNN dua dimensi telah menghasilkan hasil yang sangat menjanjikan dalam pemodelan gambar dan tekstur grayscale (Theis & Bethge, 2015). RNN dua dimensi dapat diterapkan pada pemodelan gambar alam berskala besar. Pixeln RNN yang dihasilkan terdiri dari/hingga dua belas memori *Long Short Term Memory* (LSTM) lapisan dua dimensi. Lapisan ini menggunakan unit LSTM dalam statusnya (Hochreiter & Schmidhuber, 1997; Graves & Schmidhuber, 2009) dan mengadopsi konvolusi untuk menghitung sekaligus semua status di sepanjang salah satu dimensi spasial data. Pada rancangan ini terdapat dua jenis lapisan.

Jenis pertama adalah lapisan LSTM Baris di mana konvolusi diterapkan di sepanjang setiap baris teknik serupa dijelaskan dalam (Stollenga et al., 2015). Jenis kedua adalah lapisan Diagonal BiLSTM di mana konvolusi diterapkan dengan cara baru di sepanjang diagonal gambar. Jaringan juga menggabungkan koneksi residual (He et al., 2015) di sekitar lapisan LSTM; kami mengamati bahwa ini membantu pelatihan PixelRNN hingga kedalaman dua belas lapisan.

Untuk menangkap proses pembangkitan, Theis & Bethge (2015) mengusulkan untuk menggunakan jaringan LSTM dua dimensi (Graves & Schmidhuber, 2009) yang dimulai dari piksel kiri atas dan berlanjut ke piksel kanan bawah. Keuntungan dari jaringan LSTM adalah bahwa ia secara efektif menangani ketergantungan jarak jauh yang penting bagi pemahaman objek dan pemandangan. Struktur dua dimensi memastikan bahwa sinyal disebarkan dengan baik ke arah kiri-ke-kanan dan atas-ke-bawah.

Pada bagian ini fokus pada bentuk distribusi, sedangkan bagian selanjutnya akan dikhususkan untuk mendeskripsikan inovasi arsitektur di dalam PixelRNN.



Gambar 2.8 Proses context citra

Pada gambar 2.5 dapat dijelaskan sebagi berikut:

Kiri : Untuk menghasilkan piksel xi satu kondisi pada semua piksel yang telah dibuat sebelumnya di kiri dan di atas xi.

Tengah : Untuk menghasilkan piksel dalam kasus multi-skala kita juga dapat mengkondisikan piksel gambar yang disub-sampel (dalam warna biru muda).

Kanan : Diagram konektivitas di dalam konvolusi bertopeng. Pada lapisan pertama, setiap saluran RGB terhubung ke saluran sebelumnya dan ke konteks, tetapi tidak terhubung ke saluran itu sendiri. Pada lapisan berikutnya, saluran juga terhubung ke saluran itu sendiri.

Arsitektur kedua yang disederhanakan yang memiliki komponen inti yang sama dengan Pixel RNN. *Convolutional Neural Networks* (CNN) juga dapat digunakan sebagai model sekuens dengan rentang ketergantungan tetap, dengan menggunakan konvolusi Masked. Arsitektur PixelCNN adalah jaringan konvolusional penuh dari lima belas lapisan yang mempertahankan resolusi spasial masukannya di seluruh lapisan dan keluaran distribusi bersyarat di setiap lokasi. Baik Pixel RNN dan Pixel CNN menangkap generalitas penuh dari inter-dependensi piksel tanpa memperkenalkan asumsi independen seperti misalnya model variabel laten. Ketergantungan juga dipertahankan antara nilai warna RGB dalam setiap piksel individu. Lebih lanjut, berbeda dengan pendekatan sebelumnya yang memodelkan piksel sebagai nilai berkelanjutan (misalnya, Theis & Bethge (2015); Gregor et al. (2014)), memodelkan piksel sebagai nilai diskrit

menggunakan distribusi multinomial yang diimplementasikan dengan lapisan soft-max sederhana.

2.2.4 *Bi LSTM Model*

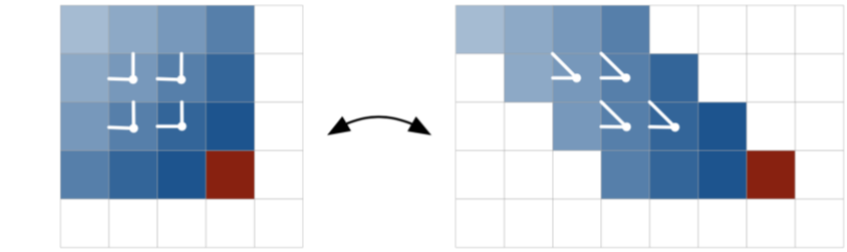
Memperkirakan distribusi gambar alami yang dapat digunakan untuk menghitung kemungkinan gambar dengan tepat dan untuk menghasilkan gambar baru. Jaringan memindai gambar satu baris pada satu waktu dan satu piksel pada satu waktu dalam setiap baris. Untuk setiap piksel, ia memprediksi distribusi bersyarat atas nilai piksel yang mungkin diberikan konteks yang dipindai, seperti yang diilustrasikan pada gambar 2.5. Distribusi gabungan atas piksel gambar difaktorkan menjadi produk distribusi bersyarat. Parameter yang digunakan dalam prediksi dibagikan ke semua posisi piksel pada gambar.

- *Generating image pixel* dengan *pixel*

Tujuannya adalah untuk menetapkan probabilitas p (x) ke setiap gambar x yang terbentuk dari n × n piksel. Kita dapat menulis gambar x sebagai urutan satu dimensi x1, ..., xn2 dimana piksel diambil dari gambar baris demi baris. Untuk memperkirakan distribusi gabungan p (x) kita menuliskannya sebagai produk dari distribusi kondisional pada piksel,

$$p(x) = \prod_{i=1}^{n^2} p(x_i | x1, ..., x_i - 1) \quad \text{.....} \quad (1)$$

Nilai p (xi | x1, ..., xi − 1) adalah probabilitas piksel ke-i xi diberikan semua piksel sebelumnya x1, ..., xi − 1. Generasi berlanjut baris demi baris dan piksel demi piksel. Gambar 2.5 (Kiri) mengilustrasikan skema pengkondisian. Setiap piksel xi pada gilirannya ditentukan bersama oleh tiga nilai.



Gambar 2.9 Diagonal BiLSTM

Dalam diagonal Bi LSTM pada gambar 2.6, untuk memungkinkan paralelisasi sepanjang diagonal, peta input dibuat miring dengan mengimbangi setiap baris dengan satu posisi sehubungan dengan baris sebelumnya. Ketika lapisan spasial dihitung dari kiri ke kanan dan kolom demi kolom, peta keluaran digeser kembali ke ukuran aslinya. Konvolusi menggunakan kernel berukuran 2×1. Satu untuk setiap saluran warna Merah, Hijau dan Biru (RGB). Distribusi p (xi | x <i) sebagai produk berikut:

$$p(x_{i,R}|x<i)p(x_{i,G}|x<i,x_{i,R})p(x_{i,B}|x<i,x_{i,R},x_{i,G}) \quad \text{.....} \quad (2)$$

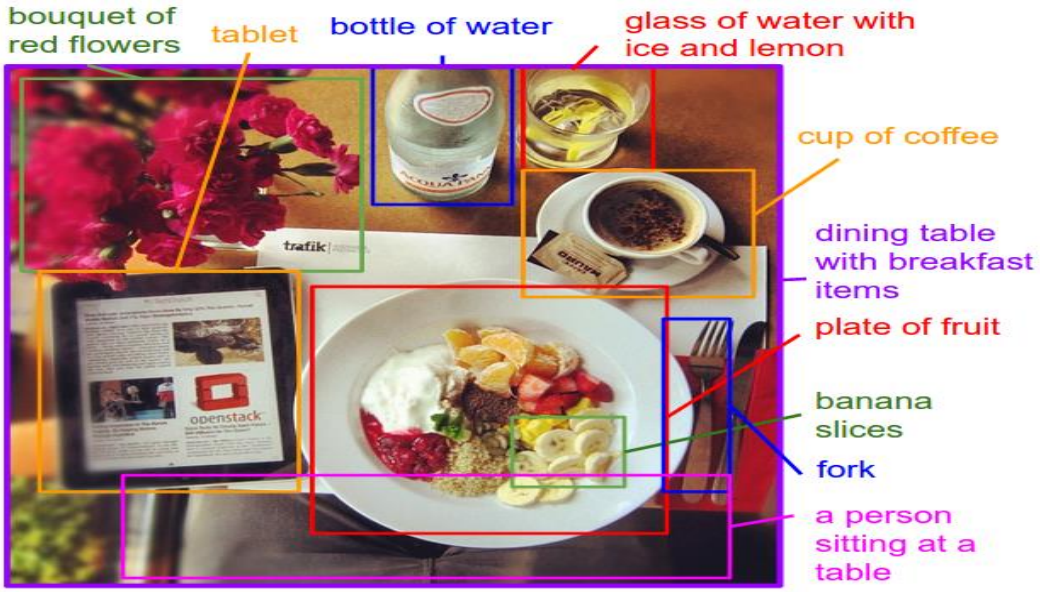
Masing-masing warna dengan demikian dikondisikan di saluran lainnya serta pada semua piksel yang dihasilkan sebelumnya. Distribusi nilai piksel dihitung secara paralel, sedangkan pembuatan gambar dilakukan secara berurutan.

BAB III

KERANGKA BERFIKIR, KONSEP PENELITIAN DAN HIPOTESIS

3.1 Kerangka Berfikir

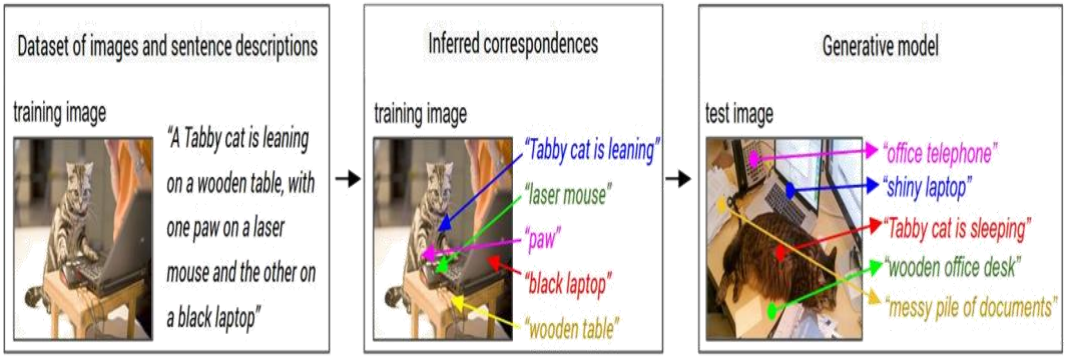
Pemahaman citra atau visual (*visual understanding*) adalah salah satu elemen inti dari konsep computer vision yang terdiri dari klasifikasi citra, deteksi objek dan segmentasi akan dilakukan melalui metode *multi task deep learning* melalui metode CNN (Convolution Neural Network) dengan teknik *Recurrent Neural Network* (RNN) *dan Long Term Short Memory* (LSTM) . Konsep ini dilakukan dengan konversi *image to text* (context) yang selanjutnya melalui proses *multi layer* untuk mengekspresikan konten gambar dalam menyelesaikan pemahaman gambar yang kompleks. Penelitian sebelumnya hanya memfokuskan pada informasi umum seperti deteksi objek beserta lokasinya, namun dalam penelitian ini akan meningkatkan pemahaman ekspresi gambar tingkat tinggi untuk menghasilkan deskripsi text dari sebuah model gambar yang merupakan konsep korespondensi antara bahasa dan visual seperti pada gambar 3.1 berikut ini. (Karpathy and Fei-Fei, 2017)



Gambar 3.1 Konsep Model Text Sebagai Ruang Label yang Menghasilkan Deskripsi pada Area Gambar

Permasalahan dari desain ini yaitu model konten gambar yang cukup kompleks untuk secara bersamaan di sinkronisasi dan direpresentasikan dalam domain bahasa. Selain itu, model harus bebas dari asumsi tentang template, aturan, atau kategori *hard-code* tertentu dan sebagai gantinya bergantung pada pembelajaran dari *training data set*. Kedua, tantangan praktisnya adalah bahwa kumpulan data teks gambar tersedia dalam jumlah besar di internet, tetapi deskripsi multipleks ini menyebutkan beberapa entitas yang lokasinya dalam gambar tidak diketahui. Model sinkronisasi ini didasarkan pada kombinasi baru *Convolutional Neural Network* (CNN) pada area gambar, dan *Recurrent Neural Network* (RNN) pada area kalimat serta tujuan terstruktur yang menyelaraskan dua modalitas melalui penyematan multimodal. Sekilas pada sebuah gambar sudah cukup bagi manusia untuk menunjukkan dan menjelaskan sejumlah besar detail tentang pemandangan visual, namun, kemampuan ini menjadi tugas yang

sulit dipahami untuk model pengenalan visual. Pada model ini proses inputnya adalah sekumpulan gambar dan deskripsi kalimat yang sesuai pada gambar 3.2 berikut ini (Karpathy and Fei-Fei, 2017).

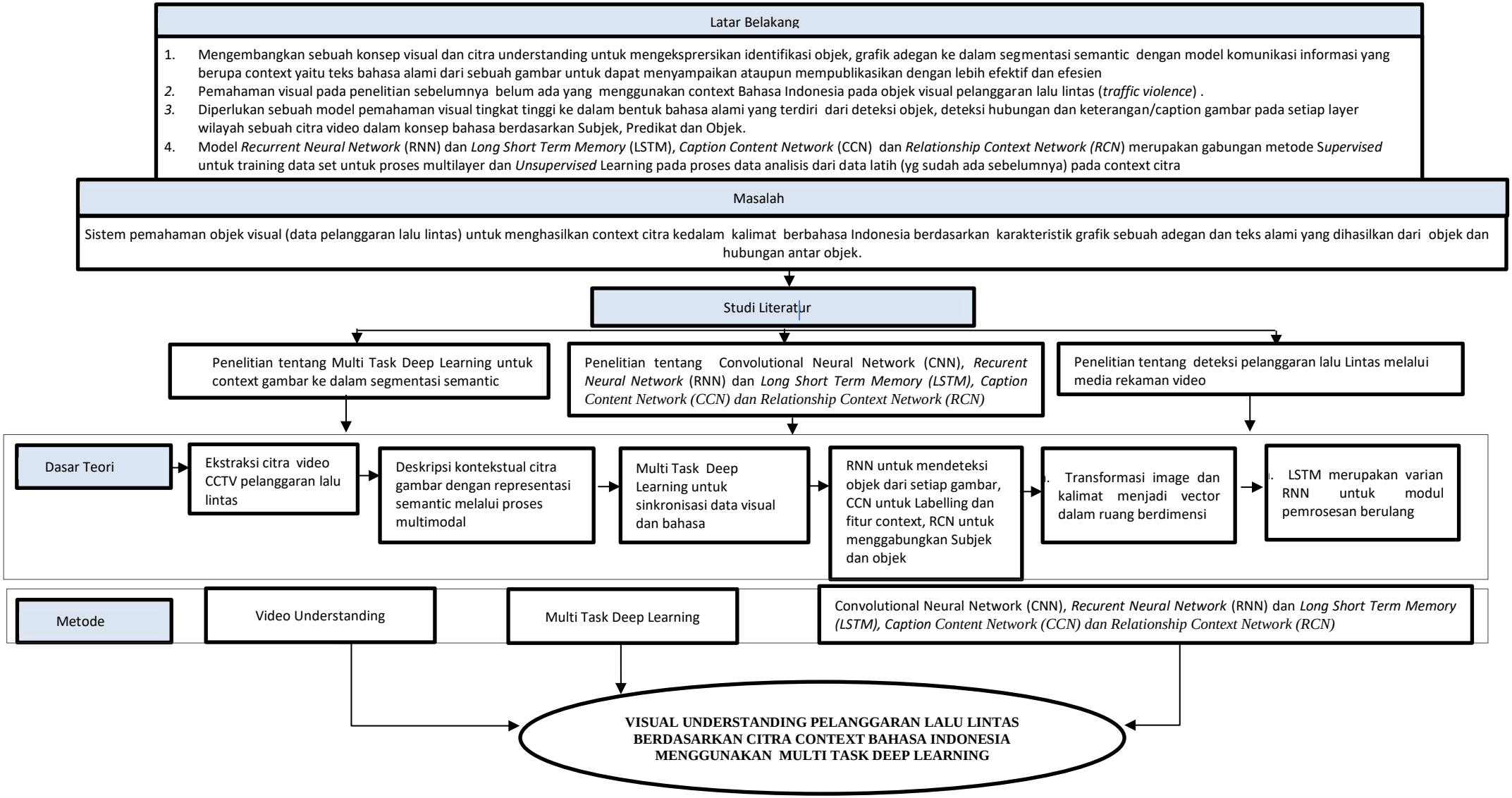


Gambar 3.2 Model Kumpulan Gambar dan Deskripsi yang Sesuai

Model di representasikan dengan menyelaraskan cuplikan kalimat ke wilayah visual yang digambarkan melalui penyematan multimodal, kemudian dikorespondensikan sebagai *training data set* untuk model RNN. Pada penelitian ini menggunakan gabungan metode *supervised learning* dengan penggunaan training data set untuk proses multilayer dan *unsupervised learning* dengan proses data analisis dilakukan tidak berdasarkan sekelompok data yang sudah ada sebelumnya (data latih) pada saat context citra gambar.

Berdasarkan hasil sintesis teori dari kajian Pustaka, serta studi literatur terhadap penelitian terdahulu, maka penelitian ini akan dibangun sistem konversi *video to text* dengan metode *multi task deep learning* dengan context citra gambar dengan Teknik *Recurrent Neural Network* dan *Long Short Term Memory* dengan fitur konversi image ke dalam bentuk text.(Mikolov *et al.*, 2010).

3.2 Konsep Penelitian



Gambar 3.3 Konsep Penelitian

BAB IV

METODE PENELITIAN

4.1 Peta Penelitian

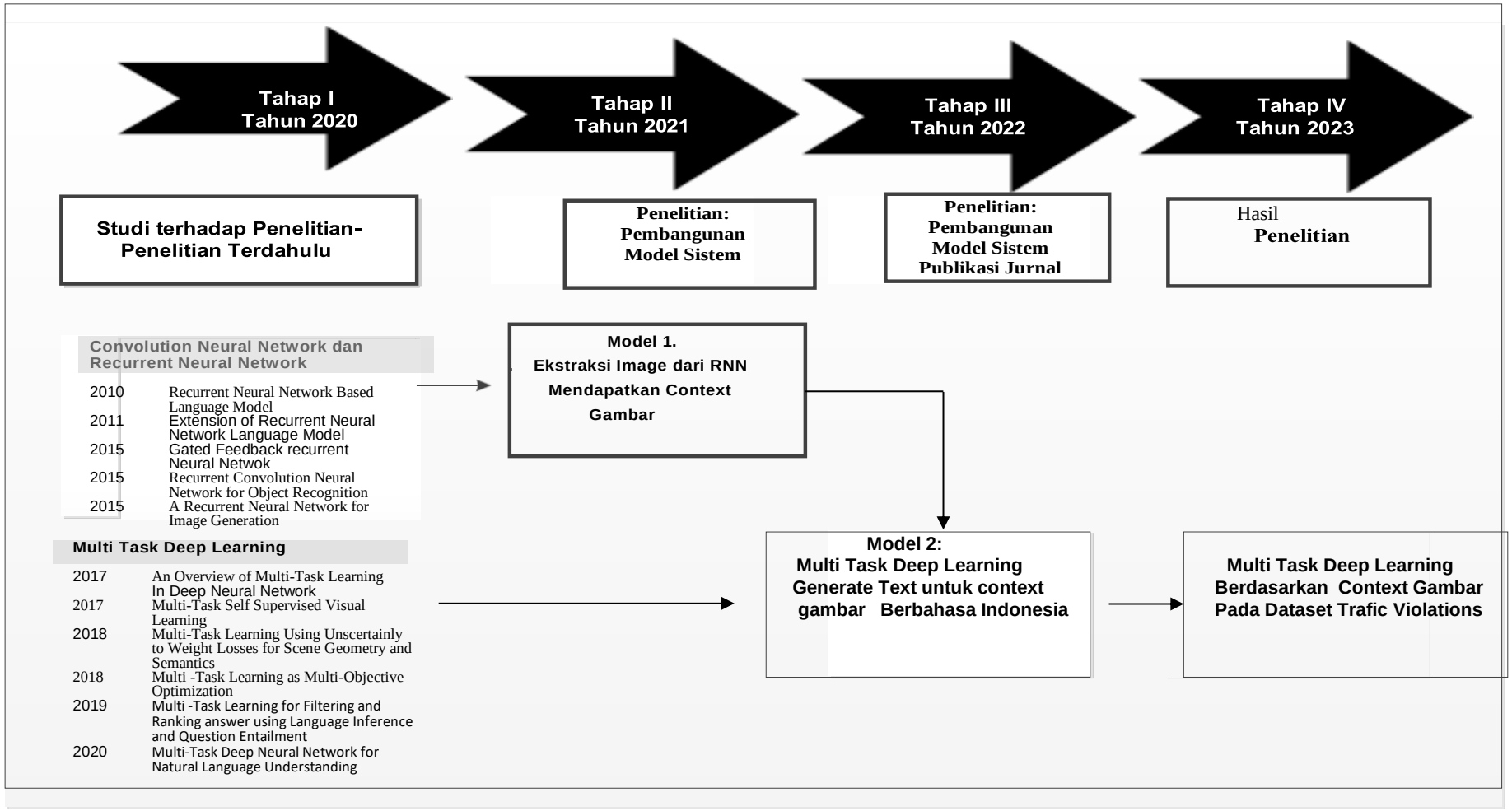
Berdasarkan pemaparan kerangka berpikir dan konsep yang dijelaskan pada Bab III, maka pembangunan sistem deteksi fake content berdasarkan context gambar menggunakan *Multi Task Deep Learning* dengan metode *RNN* dan *LSTM* dapat digambarkan dalam bentuk peta penelitian seperti pada Gambar 4.1.

4.2 Rancangan Penelitian

Pembangunan sistem deteksi *fake content* berdasarkan context gambar menggunakan *Multi Task Deep Learning* dengan metode *RNN* dan *LSTM* dapat digambarkan dalam bentuk peta penelitian seperti berikut.

- **Arsitektur Umum Model Sistem**

Berdasarkan peta penelitian, pembangunan model terbagi menjadi dua tahapan sistem yaitu Sistem 1 dan Sistem 2. Kedua sistem ini dikembangkan secara bertahap dengan melakukan perubahan dan penambahan metode. Secara umum, kedua model memiliki arsitektur yang berbeda karena sistem ini berkaitan dan harus terintegrasi satu sama lain seperti pada Gambar 4.2.



Gambar 4.1 Peta Penelitian

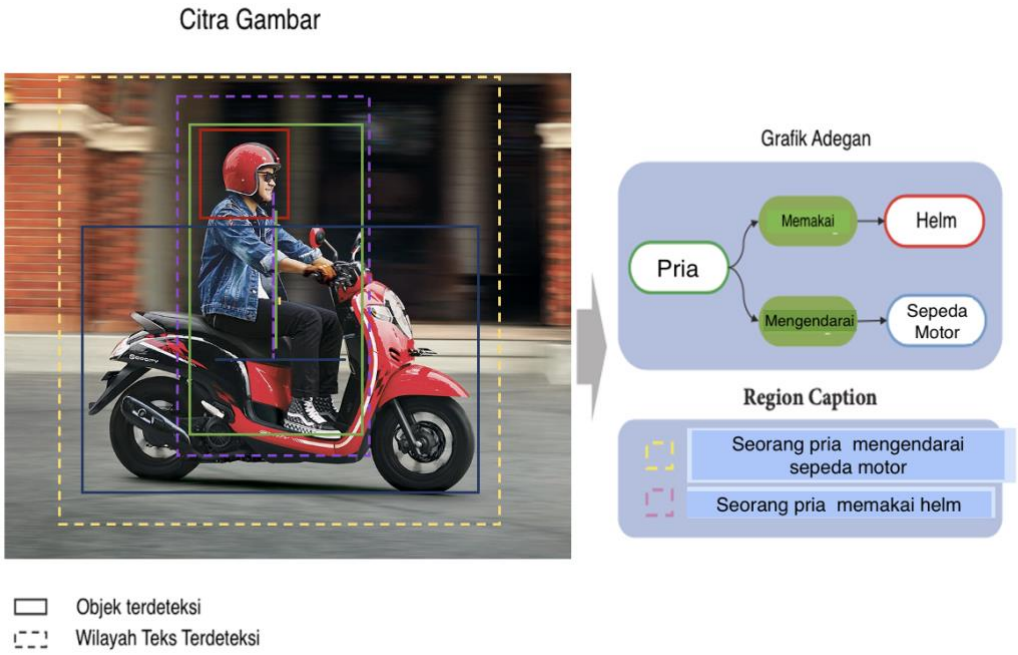
Model sistem secara garis besar dapat dibagi menjadi beberapa sub sistem/proses yaitu:

4.2.1 Sistem dan Model Arsitektur

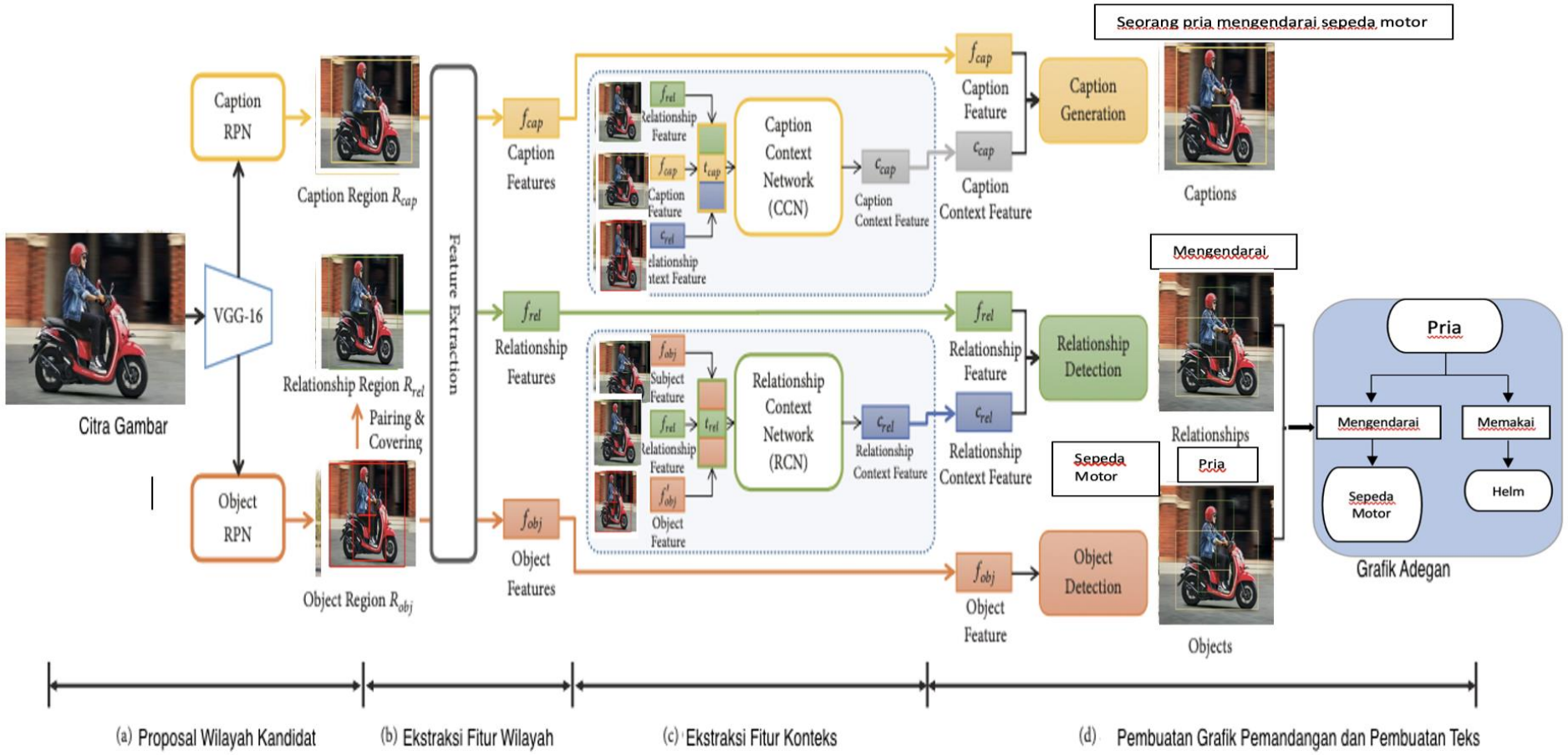
Model arsitektur pemahaman citra gambar dalam context bahasa dapat di bagi menjadi 3 proses:

1. Deteksi Objek dan wilayah teks.
2. Kerangka arsitektur sistem secara keseluruhan yang merupakan hubungan fitur objek, fitur subjek, fitur caption dan fitur context.
3. Objek dan jaringan proposal *caption region*

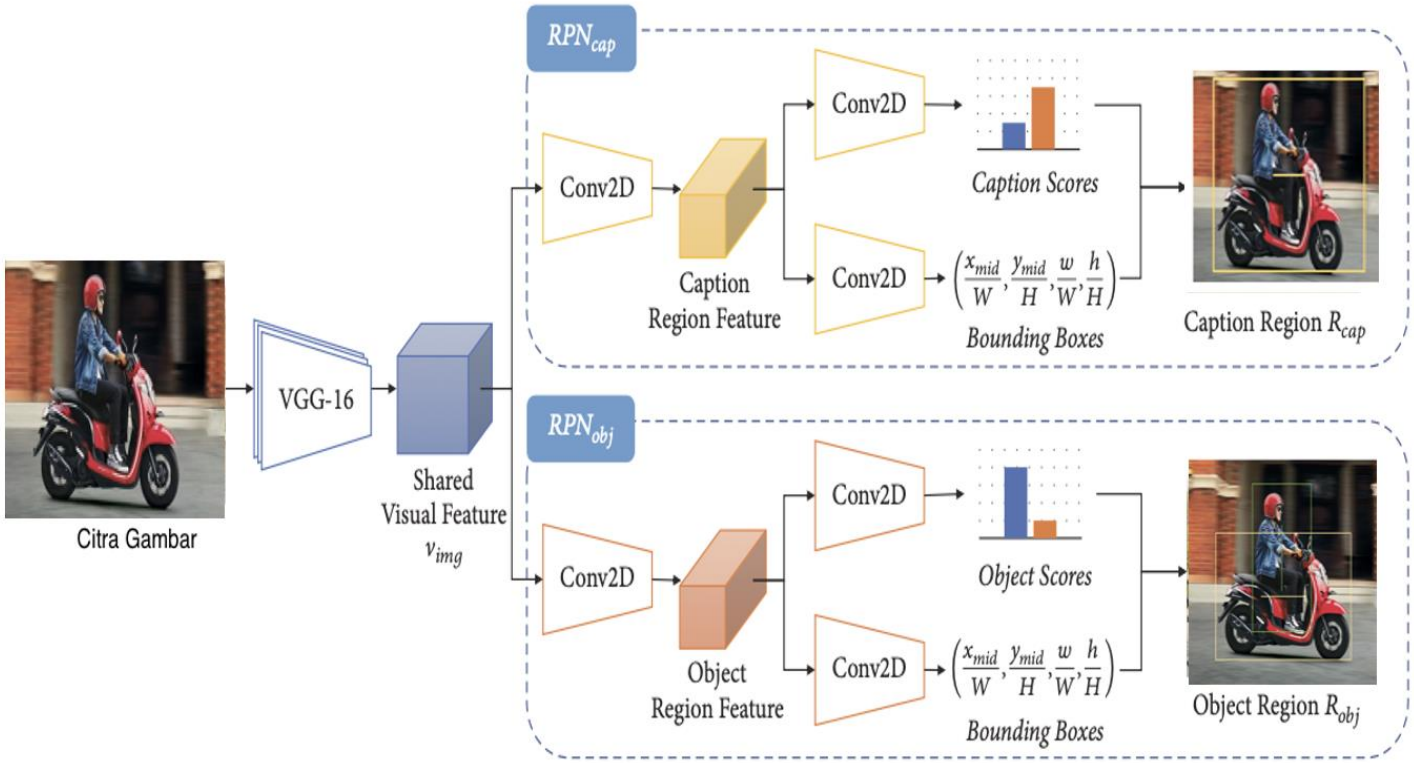
   Seperti yang di representasikan pada simulasi gambar berikut ini:



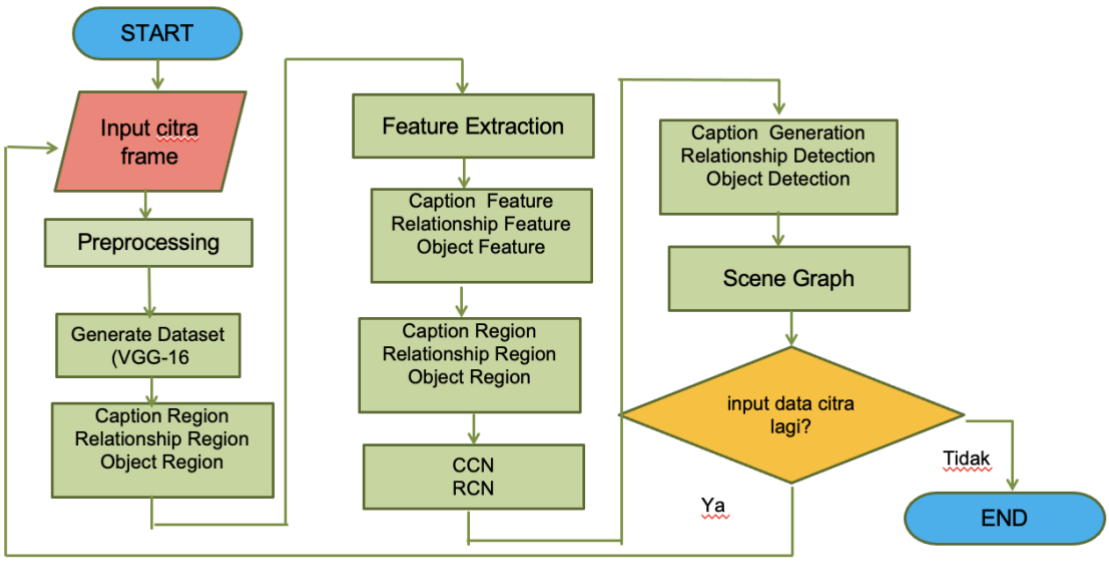
Gambar 4.2 Simulasi Pembuatan Grafik Adegan dan Pembuatan Teks Gambar



Gambar 4.3 Kerangka Keseluruhan dari Model Citra Understanding

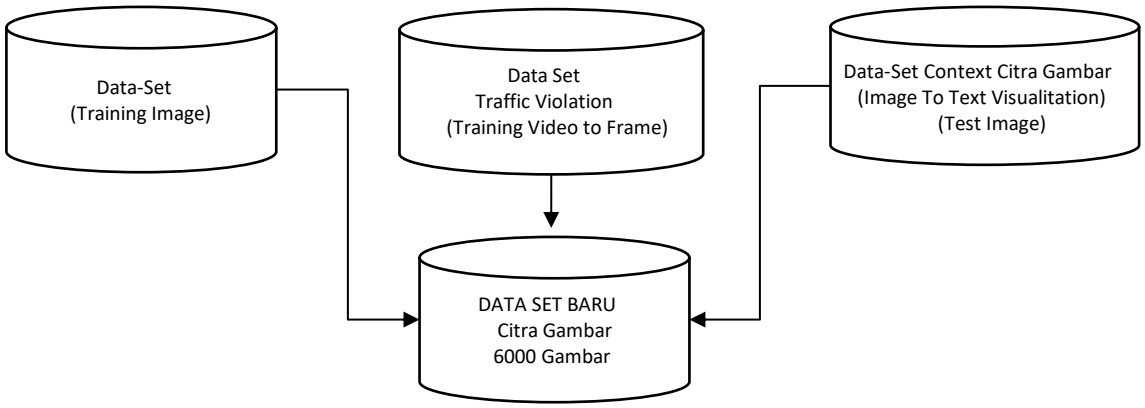


Gambar 4.4 Jaringan Deteksi Objek dan Wilayah Teks



Gambar 4.5 Alur Keseluruhan dari Model Visual Understanding

4.2.2 *Dataset Exploration*



Gambar 4.6 Eksplorasi Data-Set

4.2.3 *Preprocessing*

Pada tahapan ini terdapat empat proses yaitu

1. **Representasi Gambar**

Uraian kalimat dapat membuat sebuah referensi ke objek dan atributnya, jadi menurut metode Girshick (Karpathy and Fei-Fei, 2017) untuk mendeteksi objek di setiap gambar dengan RCNN. CNN telah dilatih sebelumnya di ImageNet dan disetel pada 200 kelas dengan menggunakan lokasi teratas yang terdeteksi selain seluruh gambar dan menghitung representasi berdasarkan piksel I_b di dalam setiap kotak pembatas sebagai berikut.

$$v = W_m[CNN_{\swarrow c}(I_b)] + b_m, \quad \text{.....} \quad (1)$$

Dimana CNN (I_b) mengubah piksel di dalam kotak pembatas I_b menjadi aktivasi dimensi 4096 dari lapisan yang terhubung sepenuhnya tepat sebelum pengklasifikasi. Parameter CNN $\surd_c$ berisi sekitar 60 juta parameter. Matriks W_m memiliki dimensi $h \twoheadrightarrow 4096$, di mana h adalah ukuran ruang embedding multimodal (h berkisar antara 1000-1600 dalam data training. Setiap gambar dengan demikian direpresentasikan sebagai himpunan vektor berdimensi-h $\{v_i \mid i = 1...20\}$. (Schuster and Paliwal, 1997).

2. Representasi Kalimat

Untuk menetapkan hubungan antar-modal, maka untuk merepresentasikan kata-kata dalam kalimat dalam ruang embedding dimensi-h yang sama dengan region gambar. Pendekatan paling sederhana mungkin dengan memproyeksikan setiap kata yang dibagi langsung ke dalam embedding ini. Namun, pendekatan ini tidak

mempertimbangkan informasi urutan dan konteks kata dalam kalimat. Untuk mengatasi masalah ini, maka digunakan Jaringan Neural Berulang Bidirec-tional atau *Bidirectional Recurrent Neural Network* (BRNN) untuk menghitung representasi kata. BRNN mengambil urutan N kata (dikodekan dalam representasi 1-of-k) dan trans-bentuk masing-masing menjadi vektor h-dimensi. Namun, representasi dari setiap kata diperkaya oleh konteks berukuran bervariasi di sekitar kata itu. Menggunakan indeks t = 1. . . N untuk menunjukkan posisi sebuah kata dalam sebuah kalimat, bentuk tepatnya dari BRNN adalah persamaan sebagai berikut:

$$x_t = W_{w\ t} \quad \text{.....} \quad (2)$$

$$e_t = f(W_e x_t + b_e) \quad \text{.....} \quad (3)$$

$$h^f = f(e_t + W_f\, h^f \quad \text{.....} \quad (4)$$

$$h_t^b = f(e_t + W_b h_t^{b}{}_{+1} + b_b) \quad \text{.....} \quad (5)$$

$$s_t = f(W_d(h_t^f + h_t^b) + b_d) \quad \text{.....} \quad (6)$$

Di sini, $_t$ adalah vektor kolom indikator yang memiliki satu di indeks kata ke-t dalam sebuah kosakata kata. Bobot W_w menentukan matriks embedding kata yang kami inisialisasi dengan bobot *word2vec* berdimensi 300 dan tetap dipertahankan karena masalah *overfitting.* Namun, dalam praktiknya kami menemukan sedikit perubahan dalam kinerja akhir saat vektor ini dilatih, bahkan dari itialisasi acak.

Perhatikan bahwa BRNN terdiri dari dua aliran pemrosesan independen, satu bergerak dari kiri ke kanan (h^f_t) dan yang lainnya dari kanan ke kiri (h^b_t). Representasi dimensi-h terakhir s_t untuk kata ke-t adalah fungsi dari kata di lokasi tersebut dan juga konteks sekitarnya dalam kalimat. Secara teknis, setiap s_t adalah fungsi dari semua kata di seluruh kalimat, representasi kata terakhir (s_t) selaras paling kuat dengan konsep visual kata di lokasi itu ($_t$). Parameter W_e, W_f, W_b, W_d dan bias masing-masing b_e, b_f, b_b, b_d. Ukuran khas dari representasi tersembunyi dalam sistem ini berkisar antara 300-600 dimensi. Fungsi aktivasi f ke unit linier yang diperbaiki (ULT), yang menghitung f: x 7! Max (0, x).(Karpathy and Fei-Fei, 2017)

3. *Allignment Objective* (Transformasi Gambar dan Kalimat Menjadi Sekumpulan Vektor Dalam Ruang Berdimensi-h yang Sama)

Karena pengawasan berada pada level gambar dan kalimat secara keseluruhan, maka dirumuskan skor gambar-kalimat sebagai fungsi dari skor kata-wilayah individu . Secara intuitif, pasangan kalimat-gambar harus memiliki skor kecocokan yang tinggi jika kata-katanya memiliki dukungan yang meyakinkan dalam gambar. Model Karpathy et (Xu *et al.*, 2015) mengartikan hasil perkalian titik $v_i^T s_t$ antara daerah ke-i dan kata ke-t sebagai ukuran kemiripan dan menggunakannya untuk menentukan skor antara gambar k dan kalimat l sebagai :

$$S_{kl} = \sum_{t2g_l} \sum_{i2g_k} \max(0, v_i^T s_t). \quad \text{..... (7)}$$

Di sini, setiap kata s_t sejajar dengan satu wilayah gambar terbaik. Seperti yang ditunjukkan dalam eksperimen, model yang disederhanakan ini juga mengarah pada peningkatan kinerja peringkat akhir. Dengan asumsi bahwa k = l menunjukkan gambar dan pasangan kalimat yang sesuai, margin maksimum akhir, kerugian terstruktur tetap ada.

$$S_{kl} = \sum_{t2g_l} \max_{i2g_k} v_i^T s_t . \quad \text{..... (8)}$$

$$C(\checkmark) = \sum_k h \Big|_{\text{rank}} \{z_{\text{images}}\} \max(0, S_{kl} - S_{kk} + 1) + \sum_i \Big|_{\text{rank}} \{z_{\text{sentences}}\} \max(0, S_{lk} - S_{kk} + 1) . \quad \text{..... (9)}$$

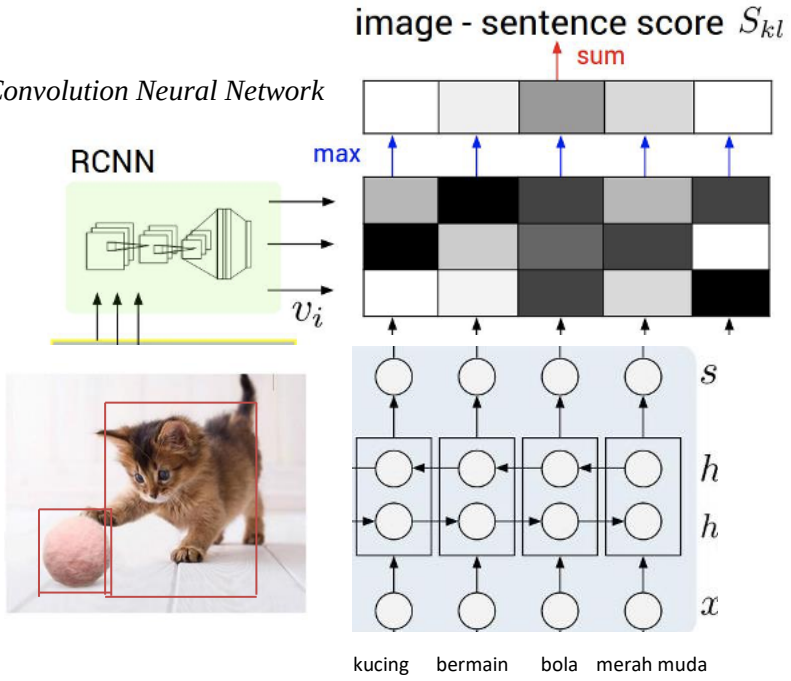
Tujuan ini mendorong pasangan gambar-kalimat yang selaras untuk memiliki skor yang lebih tinggi daripada pasangan yang tidak selaras, dengan margin.

4. *Decoding Text Segment Alignments to Image* (Mendekodekan Segment Text ke Gambar)

Proses ini dilakukan dengan menyelaraskan gambar dari set pelatihan dan kalimat yang sesuai. Kita dapat menafsirkan kuantitas $v_i^T s_t$ sebagai probabilitas log yang tidak dinormalisasi dari kata ke-t

yang menggambarkan kotak pembatas pada gambar. Namun, pada akhirnya untuk membuat cuplikan teks daripada satu kata, dengan menyelaraskan urutan kata yang diperpanjang dan bersebelahan ke satu kotak pembatas. Perhatikan bahwa solusi naïıve yang menetapkan setiap kata secara terpisah ke wilayah dengan skor tertinggi tidak cukup karena menyebabkan kata-kata tersebar secara tidak konsisten ke wilayah yang berbeda

Recurrent Convolution Neural Network



Gambar 4.7
Diagram Evaluasi citra-kalimat dalam Skor S_{kl} dengan contoh objek.

Wilayah objek disematkan dengan CNN (kiri). Kata-kata (diperkaya oleh konteksnya) disematkan dalam ruang multimodal yang sama dengan BRNN (kanan). Kemiripan berpasangan dihitung dengan produk dalam (besaran ditampilkan dalam skala abu-abu) dan akhirnya dikurangi menjadi skor kalimat gambar dengan persamaan 8 .

keselarasan ke wilayah yang sama. Secara konkret, diberikan kalimat dengan N kata dan gambar dengan kotak pembatas M, kami memperkenalkan variabel perataan laten $a_j 2 \{1...M\}$ untuk $j = 1...N$ dan rumuskan MRF dalam struktur cha di sepanjang kalimat sebagai berikut:

$$E(a) = \sum_{j=1...N} {}_j^U(a_j) + \sum_{j=1...N-1} {}_j^B(a_j, a_{j+1}) \quad (10) \qquad \ldots\ldots\ldots\ldots (10)$$

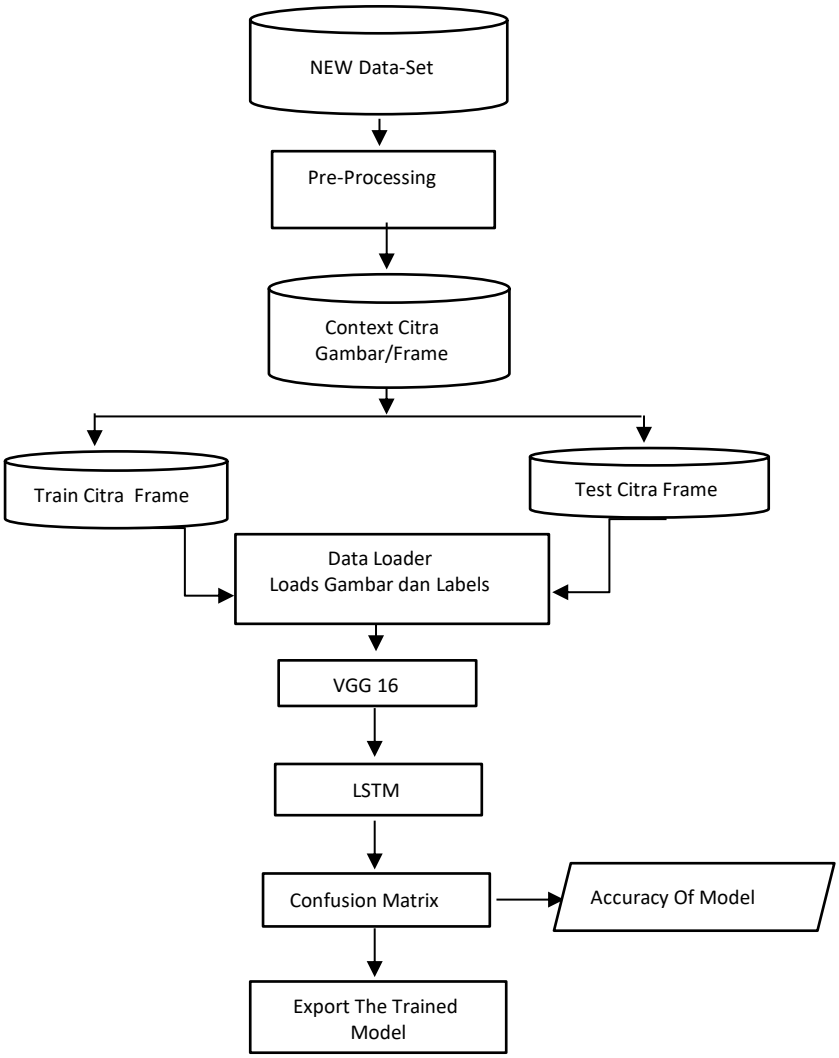
$${}_j^U(a_j = t) = v_i^T s_t \qquad \ldots\ldots\ldots\ldots (11)$$

$${}_j^B(a_j, a_{j+1}) = \beta [a_j = a_{j+1}] \qquad \ldots\ldots\ldots\ldots (12)$$

Di sini, β adalah hyperparameter yang mengontrol afinitas ke frasa kata yang lebih panjang. Parameter ini memungkinkan kita untuk melakukan interpolasi antara perataan satu kata (β = 0) dan menyelaraskan seluruh kalimat ke satu, wilayah skor maksimal ketika β besar. Hal ini dapat meminimalkan energi untuk menemukan penyelarasan terbaik dengan menggunakan pemrograman dinamis. Keluaran dari proses ini adalah sekumpulan wilayah gambar yang dianotasi dengan segmen teks.

4.2.4 *Training Workflow*

Pada diagram alur data latih berikut ini di gambarkan proses context gambar dari data set setelah proses context citra gambar sebelumnya.



Gambar 4.8 Training Workflow

4.3 Teknik Pengujian

1. Gambaran hasil ekstraksi dari CNN akan mendapatkan context dari kalimat di kemudian baru dimasukkan ke dalam RNN untuk mendapatkan context dari gambar berupa *generate text* berbahasa Indonesia.
2. Untuk image dari fitur CNN nanti langsung masuk ke dalam *multitask learning* sete bersama dengan klasifikasi text akan di hitung scorenya

bersama dengan hasilnya adalah teks caption dari gambar berbahasa Indonesia.

3. Pengujian hasil konversi gambar ke text berbahasa Indonesia berbasis Multi Task Deep Learning dengan metode RNN menggunakan LSTM atau BERT dengan *scooring* menggunakan *f1-score, precision* dan *recall*. Hasilnya nantinya akan di plot menggunakan grafik AUC (*Area Under The Curve*) dan ROC (*Receivers Operating Characteristics*). (Ren et al., 2017).

4.4 Lokasi dan Waktu Penelitian

Penelitian yang dilakukan untuk membangun sistem deteksi konten hoax berdasarkan context ctra gambar dengan menggunakan Multi Task deep Learning menghasilkan keluaran dalam bentuk aplikasi berbasis komputer. Untuk itu, kegiatan penelitian yang dilakukan tidak berbentuk eksperimen di laboratorium, melainkan pembuatan kode- kode program yang dilakukan secara mandiri.

Penelitian ini dilakukan sebagai prasyarat dalam menyelesaikan Program Studi Doktor, sehingga waktu yang dibutuhkan untuk melakukan penelitian disesuaikan dengan masa studi perkuliahan yaitu selama tiga tahun, terhitung mulai dari semester I sampai dengan semester VI.

4.5 Instrumen Penelitian

Pembangunan sistem deteksi konten hoax berdasarkan context citra gambar dengan menggunakan *Multi Task Deep Learning* dalam penelitian ini melibatkan dua instrument penelitian, sebagai berikut ini.

1. Perangkat Keras

Perangkat keras yang digunakan dalam membangun sistem dengan spesifikasi minimal sebagai berikut:

a. Prosesor Intel Core i5-5200U Processor 2.5 GHz 3M cache
b. Memori 16 GB DDR3
c. Kartu Grafis (VGA) NVIDIA GeForce 840M 2 GB
d. Harddisk 500 GB

2. Perangkat Lunak

Perangkat lunak yang digunakan dalam membangun sistem sebagai berikut:

a. MacOS Big Sur Version 14.4
b. Bahasa Pemrograman Python, dengan repositori dan banyak fungsi paket diantaranya *automation*, analisis data, database, dokumentasi antar muka pengguna grafis, pemgolahan citra, *machine learning*, multimedia *text processing*, *web scraping* dan lain sebagainya.
c. *Library Phyton* yaitu TensorFlow, untuk menghitung komputasi numerik secara mudah dan cepat.
d. Xampp untuk menjalankan PHP, MySQL dan Apache dalam server lokal.

4.6 Prosedur Penelitian

Penelitian yang dilakukan untuk membangun sistem deteksi konten hoax berdasarkan context ctra gambar dengan menggunakan Multi Task Deep Learning dalam penelitian ini, mengikuti prosedur yang ditetapkan dalam Siklus Hidup Pengembangan Perangkat Lunak seperti yang digambarkan pada Gambar 4.7

Berdasarkan Gambar 4.7 prosedur dalam penelitian ini dapat dijelaskan sebagai berikut :

1. **Tahap Analisis**

   Pada tahap ini, dilakukan penentuan kebutuhan terhadap model sistem yang akan dikembangkan, meliputi data yang terlibat dalam sistem baik masukan maupun keluaran, pengguna sistem, serta proses-proses yang dibutuhkan untuk mengolah data.

2. **Tahap Desain**

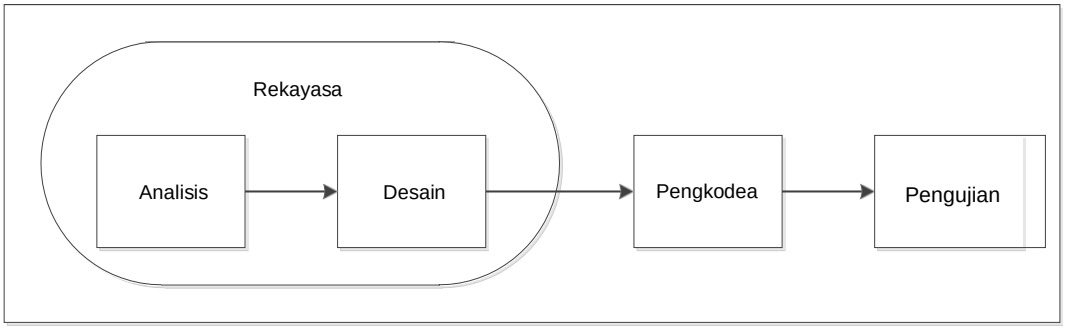
   Pada tahap ini, dilakukan proses perancangan model sistem, meliputi pendefinisian dan pembuatan basis data, perancangan algoritma, dan perancangan antarmuka sebagai media interaksi antara pengguna dan sistem.

3. **Tahap Pengkodean**

   Pada tahap ini, dilakukan proses pengkodean rancangan algoritma menggunakan instrumen perangkat keras dan perangkat lunak, sehingga menghasilkan sistem yang siap untuk diuji.

4. **Tahap Pengujian**

   Pada tahap ini, dilakukan pengujian terhadap sistem, untuk menentukan apakah sistem perlu diperbaiki, atau siap untuk diimplementasikan. (Setiawan, 2015).



Gambar 4.9 Siklus Hidup Pengembangan Sistem

Sumber: Software Engineering: A Practitioner's Approach pada Pressman (2010)

DAFTAR PUSTAKA

Aini, Z., Hutapea, F. and Ramadhanie, N. (2020) "Di Kota Tanjungpinang (Studi Kasus Dinas Perhubungan)," 11, pp. 1–13.

Alkan, B. *et al.* (2019) "Driver cell phone usage violation detection using license plate recognition camera images," *VEHITS 2019 - Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems*, (Vehits), pp. 468–474. doi: 10.5220/0007725804680474.

Ariyoga, D., Rahmadi, R. and Rajagede, R. A. (2021) "Penelitian Terkini Tentang Sistem Pendeteksi Pelanggaran Lalu Lintas Berbasis Deep Learning : Sebuah Kajian Pustaka," *Automata*, 2(1).

Bai, S. *et al.* (2018) "Natural language guided visual relationship detection," *Mathematical Problems in Engineering*, 2020(1), pp. 444–453. doi: 10.1145/3219819.3220036.

Bai, S. and An, S. (2018) "A survey on automatic image caption generation," *Neurocomputing*, 311, pp. 291–304. doi: 10.1016/j.neucom.2018.05.080.

Bouwmans, T. *et al.* (2015) "Recent Advanced Statistical Background Modeling for Foreground Detection - A Systematic Survey Thierry Bouwmans To cite this version : Recent Advanced Statistical Background Modeling for Foreground Detection - A Systematic Survey."

Bregler, C., Covell, M. and Slaney, M. (1997) "Video Rewrite: Driving visual speech with audio," *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1997*, (November 2014), pp. 353–360. doi: 10.1145/258734.258880.

Guo, Y. *et al.* (2016) "Deep learning for visual understanding: A review," *Neurocomputing*, 187, pp. 27–48. doi: 10.1016/j.neucom.2015.09.116.

Jácome-Galarza, L. R. *et al.* (2020) "Computer Vision for Image Understanding: A Comprehensive Review," *Advances in Intelligent Systems and Computing*, 1066(May), pp. 248–259. doi: 10.1007/978-3-030-32022-5_24.

Jeeva, S. and Sivabalakrishnan, M. (2015) "Survey on background modeling and

foreground detection for real time video surveillance," *Procedia Computer Science*, 50, pp. 566–571. doi: 10.1016/j.procs.2015.04.085.

K. Klubsuwan, W. K. and S. M. (2013) "Traffic Violation Detection Using Multiple Trajectories Evaluation of Vehicles," *4th International Conference on Intelligent Systems,* 10.1109/IS, pp. 220-224,. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6498268&isnumber=6498213.

Karpathy, A. and Fei-Fei, L. (2017) "Deep Visual-Semantic Alignments for Generating Image Descriptions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), pp. 664–676. doi: 10.1109/TPAMI.2016.2598339.

Kinghorn, P., Zhang, L. and Shao, L. (2018a) "A region-based image caption generator with refined descriptions," *Neurocomputing*, 272, pp. 416–424. doi: 10.1016/j.neucom.2017.07.014.

Kinghorn, P., Zhang, L. and Shao, L. (2018b) "A region-based image caption generator with refined descriptions," *Neurocomputing*, 272, pp. 416–424. doi: 10.1016/j.neucom.2017.07.014.

Liu, X. *et al.* (2019) "Multi-task deep neural networks for natural language understanding," *arXiv*, pp. 4487–4496.

Mikolov, T. *et al.* (2010) "Recurrent neural network based language model," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010*.

Mnih, V. *et al.* (2014) "Recurrent models of visual attention," in *Advances in Neural Information Processing Systems*.

Van Den Oord, A., Kalchbrenner, N. and Kavukcuoglu, K. (2016) "Pixel recurrent neural networks," *33rd International Conference on Machine Learning, ICML 2016*, 4, pp. 2611–2620.

Putra, D. (2010) "Pengolahan Citra Digital," (April), p. 420.

Ren, S. *et al.* (2017) "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), pp. 1137–1149. doi:

10.1109/TPAMI.2016.2577031.

Ruder, S. (2017) “An Overview of Multi-Task Learning in Deep Neural Networks∗,” *arXiv*.

Sarraf, A., Azhdari, M. and Sarraf, S. (2021) “A Comprehensive Review of Deep Learning Architectures for Computer Vision Applications,” *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 77(1), pp. 1–29.

Schuster, M. and Paliwal, K. K. (1997) “Bidirectional recurrent neural networks,” *IEEE Transactions on Signal Processing*. doi: 10.1109/78.650093.

Setiawan, A. (2015) “Implementasi Aplikasi Decision Support System dengan Metode Analytical Hierarcy Process (AHP) untuk Penentuan Jenis Supplier,” *Jurnal Gaung Informatika*, pp. 1–10.

Shin, D. and Kim, I. (2018) “Deep Image Understanding Using Multilayered Contexts,” *Mathematical Problems in Engineering*. Edited by A. A. Salman, 2018, p. 5847460. doi: 10.1155/2018/5847460.

Singh, V., Unadkat, V. and Kanani, P. (2019) “Intelligent traffic management system,” *International Journal of Recent Technology and Engineering*, 8(3), pp. 7592–7597. doi: 10.35940/ijrte.C6168.098319.

Staniute, R. and Šešok, D. (2019) “A systematic literature review on image captioning,” *Applied Sciences (Switzerland)*, 9(10). doi: 10.3390/app9102024.

Umam, K. and Negara, B. S. (2016) “Deteksi Obyek Manusia Pada Basis Data Video Menggunakan Metode Background Subtraction Dan Operasi Morfologi,” *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 2(2), p. 31. doi: 10.24014/coreit.v2i2.2391.

Wang, P. *et al.* (2019) “A single-shot arbitrarily-shaped text detector based on context attended multi-task learning,” *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, (1), pp. 1277–1285. doi: 10.1145/3343031.3350988.

Wiriyathammabhum, P. *et al.* (2017) “Computer Vision and Natural Language Processing,” *ACM Computing Surveys*, 49(4), pp. 1–44. doi:

10.1145/3009906.

Xu, K. *et al.* (2015) “Show, attend and tell: Neural image caption generation with visual attention,” *32nd International Conference on Machine Learning, ICML 2015*, 3, pp. 2048–2057.

Z.Zang (2010) “Research on the taxi traffic accident and violation identification model,” *IEEE XPlore*, pp. 533–536.