

TESIS

**PENDETEKSIAN *DEEFAKE* PADA VIDEO
MENGUNAKAN RESNEXT CNN DAN *LONG SHORT-TERM
MEMORY***

*Deepfake Detection On Video Using ResNext CNN And Long Short-
Term Memory*

**MUHAMMAD INDRA ABIDIN
D032181008**



**PROGRAM STUDI MAGISTER TEKNIK ELEKTRO
DEPARTEMEN TEKNIK ELEKTRO
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2023**

PENGAJUAN TESIS

**PENDETEKSIAN *DEEFAKE* PADA VIDEO
MENGUNAKAN RESNEXT CNN DAN *LONG SHORT-TERM
MEMORY***

Tesis
Sebagai Salah Satu Syarat Untuk Mencapai Gelar Magister
Program Studi Teknik Elektro

Disusun dan diajukan oleh

**MUHAMMAD INDRA ABIDIN
D032181008**

Kepada

**FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2023**

TESIS

**PENDETEKSIAN *DEEFAKE* PADA VIDEO
MENGUNAKAN RESNEXT CNN DAN *LONG SHOT-
TERM MEMORY***

MUHAMMAD INDRA ABIDIN
D032 181 008

Telah dipertahankan di hadapan Panitia Ujian Tesis yang dibentuk dalam rangka penyelesaian studi pada Program Magister Teknik Elektro, Fakultas Teknik

Universitas Hasanuddin

pada tanggal 03 Januari 2023

dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama



Dr. Ir. Ingrid Nurtanio, M.T
NIP. 19610813 198811 2 001

Pembimbing Pendamping



Prof. Dr. Ir. H. Andani Achmad, M.T
NIP. 19601231 198703 1 022

Dekan Fakultas Teknik
Universitas Hasanuddin



Prof. Dr. Eng. Ir. Muhammad Isran Ramli, S.T., M.T
NIP. 19730926 2000 121 002

Ketua Program Studi
S2 Teknik Elektro



Dr. Eng. Ir. Wardi, S.T., M.Eng
NIP. 19720828 199903 1 003

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Yang bertanda tangan di bawah ini

Nama : Muhammad Indra Abidin
Nomor mahasiswa : D032181008
Program Studi : Teknik Elektro

Dengan ini menyatakan bahwa, tesis berjudul “Pendeteksian *Deepfake* Pada Video Menggunakan ResNext CNN Dan *Long Short-Term Memory*” adalah benar karya saya dengan arahan dan komisi pembimbing (Dr. Ir. Ingrid Nurtanio, M.T. dan Prof. Dr. Ir. Andani Achmad, M.T.). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal/Prosiding (ILKOM Jurnal Ilmiah, Vol. 14 No. 3, dan <https://doi.org/10.33096/ilkom.v14i3.1254.178-185>) sebagai artikel dengan judul “*Deepfake Detection in Videos Using Long Short-Term Memory and CNN ResNext*”.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 03 Januari 2023

Yang menyatakan,



Muhammad Indra Abidin

KATA PENGANTAR

Segala puji selalu dipanjatkan kepada Allah SWT Yang Maha Kuasa yang telah memberikan rahmat, hidayah dan pertolongan-Nya dalam menyelesaikan tesis, yang berjudul **“Pendeteksian *Deepfake* pada Video Menggunakan ResNext CNN dan *Long Short-Term Memory*.”** Tak lupa pula shalawat dan salam kepada Nabi Muhammad SAW yang telah menyinari dunia ini dengan keindahan ilmu dan akhlak yang diajarkan kepada seluruh umatnya.

Ucapan terima kasih pun penulis haturkan kepada kepada dosen pembimbing tesis Dr. Ir. Ingrid Nurtanio, M.T. dan Prof. Dr. Ir. Andani Achmad, M.T. yang telah meluangkan waktunya untuk membimbing dan berkonsultasi tentang materi dalam tesis ini dan juga kepada seluruh dosen dan staf Departemen Teknik Elektro, Universitas Hasanuddin yang telah membantu dalam hal keilmuan maupun administrasi pada tahap tesis ini. Penulis menyadari bahwa tesis ini masih belum sempurna. Dengan demikian, penulis tetap mengharapkan kritik dan saran dengan harapan tulisan ini bisa memberikan manfaat kepada seluruh pihak.

Makassar, Januari 2023

Penulis

ABSTRAK

MUHAMMAD INDRA ABIDIN. *Pendeteksian Deepfake pada Video Menggunakan ResNext CNN dan Long Short-Term Memory* (dibimbing oleh **Ingrid Nurtanio and Andani Achmad**).

Computer vision merupakan ilmu yang mengarah pada proses pengenalan terhadap suatu objek yang dilakukan dengan sistem komputer. Proses pengenalan yang dilakukan dapat mengatasi berbagai masalah, salah satunya adalah pemalsuan dalam video atau biasa disebut *deepfake*. *Deepfake* memungkinkan untuk membuat seseorang mengatakan atau melakukan sesuatu yang tidak pernah mereka katakan atau lakukan dalam sebuah video bahkan hingga pada pemalsuan wajah yang dapat mempengaruhi privasi. Kemudahan dalam membuat video *deepfake* menunjukkan bahwa kebutuhan metode otentikasi yang semakin meningkat untuk menklasifikasi video *deepfake*. Penelitian ini bertujuan untuk mendeteksi *deep-fake* pada video dengan menggunakan metode ResNext *Convolutional Neural Networks* (CNN) dan *Long Short-Term Memory* (LSTM). Data yang telah dikumpulkan dalam bentuk video sebanyak 200 data, dilakukan proses *preprocessing* dengan *cropping* gambar dari video menjadi 112 x 112 *pixels* dan mengelompokkan menjadi 10 *frames*, 20 *frames*, 40 *frames* dan 60 *frames*. Data yang telah melalui proses *preprocessing* akan diinputkan ke dalam sistem yang dibangun dengan metode ResNext CNN dan LSTM. Metode ResNext CNN akan melakukan *extraction* dari objek, Metode LSTM memiliki blok *memory* yang akan berfungsi untuk mengingat struktur keluaran dari data training. Kombinasi proses metode ResNext CNN dan LSTM ini akan mengambil keputusan *fake* atau *realnya* suatu objek. Kinerja sistem akan diukur dengan metode *confusion matrix*. Proses implementasi memperoleh hasil akurasi tertinggi yaitu 90% pada data video yang memiliki 60 *frames*. Sedangkan data dengan 10 *frames* memiliki akurasi terendah yaitu 52%. Pada ukuran citra yang kecil ResNext CNN-LSTM mampu mendeteksi *deepfake* pada video dengan baik.

Kata kunci: *Deepfake*, ResNext CNN, LSTM

ABSTRACT

MUHAMMAD INDRA ABIDIN. *Deepfake Detection on Video Using ResNext CNN and Long Short-Term Memory* (Supervised by **Ingrid Nurtanio** and **Andani Achmad**)

Computer vision is a science related to the process of identifying an object that is done with a computer system. One problem that requires an identification process is the Deepfake. Deepfake is a forgery using video that makes it possible to make someone say or do something they never said or did in a video even to fake faces that can affect privacy. The ease of making deepfake videos shows that the need for authentication methods is increasing to classify deepfake videos. This study aims to detect deep-fakes in videos using the ResNext Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) methods. Data that has been collected in video generates as much as 200 data, preprocessing is carried out by cropping the images from the video to 112 x 112 pixels and grouping them into 10 frames, 20 frames, 40 frames and 60 frames. The pre-processed data subsequently used as input into the system built using the ResNext CNN and LSTM methods. The CNN-ResNext function is to extract the object, while the LSTM method has a memory block which record the output structure of the training data. The combination of the CNN and LSTM ResNext method processes will make a fake or real decision on an object. System performance will be measured by the confusion matrix method. The implementation process obtains the highest accuracy results, namely 90% on video data which has 60 frames. While data with 10 frames has the lowest accuracy of 52%. At a small image size, ResNext CNN-LSTM is able to detect deepfakes in videos well.

Keywords: Deepfake, ResNext CNN, LSTM.

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
PENGAJUAN TESIS.....	ii
PERSETUJUAN TESIS	iii
PERNYATAAN KEASLIAN TESIS	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Batasan Penelitian	4
1.6 Sistematika Penulisan.....	4
BAB II	6
TINJAUAN PUSTAKA	6
2.1 <i>Image Processing</i>	6
2.1.1 <i>Image acquisition</i>	6
2.1.2 <i>Pre-processing</i>	6
2.1.3 <i>Feature extraction</i>	7
2.2 <i>Long Short Term Memory (LSTM)</i>	7
2.3 <i>ResNext Convolutional Neural Network (ResNext CNN)</i>	10
2.4 Evaluasi Proses Pendeteksian	12
2.5 <i>State of the Art</i>	14
2.6 Kerangka Pikir	15
BAB III.....	16

METODE PENELITIAN	16
3.1 Tahapan Penelitian	16
3.2 Waktu dan Lokasi Penelitian	17
3.3 Jenis Penelitian.....	17
3.4 Proses Pengumpulan Data.....	17
3.5 Perancangan Sistem	18
3.6 Instrumentasi Penelitian.....	26
BAB IV	27
HASIL DAN PEMBAHASAN	27
4.1 Hasil Penelitian	27
4.1.1 Hasil <i>feature extraction</i> dengan metode ResNext CNN	27
4.1.2 Hasil klasifikasi dengan metode LSTM.....	28
4.1.3 Hasil pengujian.....	28
4.2 Pembahasan.....	35
4.2.1 Pembahasan hasil implementasi sistem	35
4.2.2 Analisis kinerja sistem	40
BAB V.....	42
KESIMPULAN DAN SARAN	42
5.1 Kesimpulan	42
5.2 Saran.....	42
DAFTAR PUSTAKA	43
LAMPIRAN.....	46

DAFTAR TABEL

Nomor	Halaman
Tabel 1 <i>Confusion matrix</i>	13
Tabel 2 <i>State of the Art</i>	14
Tabel 3 <i>Feature extraction</i> dengan ResNext CNN	23
Tabel 4 Instrumentasi perangkat penelitian.....	26
Tabel 5 Hasil deteksi <i>deepfake</i> dengan metode ResNext CNN dan LSTM.....	29
Tabel 6 Pengujian pada 100 <i>dataset 10 frames</i>	30
Tabel 7 Pengujian pada 100 <i>dataset 20 frames</i>	31
Tabel 8 Pengujian pada 100 <i>dataset 40 frames</i>	32
Tabel 9 Pengujian pada 100 <i>dataset 60 frames</i>	33
Tabel 10 Performa sistem dalam pengujian <i>confusion matrix</i>	39

DAFTAR GAMBAR

Nomor	Halaman
Gambar 1 <i>Pseudo-Code</i> klasifikasi metode LSTM (Nakisa, Rastgoo, Rakotonirainy, Maire, & Chandran, 2018)	7
Gambar 2 Detail struktur LSTM	8
Gambar 3 Struktur <i>forget layer</i>	8
Gambar 4 Struktur <i>remember gate</i>	9
Gambar 5 Struktur <i>update layer</i>	9
Gambar 6 Struktur <i>output layer</i>	10
Gambar 7 Arsitektur MLP sederhana.....	11
Gambar 8 Ilustrasi proses konvolusi	11
Gambar 9 Proses konvolusi pada ResNext CNN	12
Gambar 10 Kerangka pikir	15
Gambar 11 Tahapan penelitian.....	16
Gambar 12 Perancangan sistem.....	18
Gambar 13 Hasil <i>splitting</i> video dalam <i>frame</i>	19
Gambar 14 Hasil <i>face detection</i>	20
Gambar 15 Proses <i>cropping</i>	20
Gambar 16 Hasil tahapan <i>pre-processing</i> 10 <i>frames</i>	21
Gambar 17 Hasil tahapan <i>pre-processing</i> 20 <i>frames</i>	21
Gambar 18 Hasil tahapan <i>pre-processing</i> 40 <i>frames</i>	22
Gambar 19 Hasil tahapan <i>pre-processing</i> 60 <i>frames</i>	22
Gambar 20 Tahapan metode ResNext CNN.....	23
Gambar 21 Kode program klasifikasi metode LSTM	24
Gambar 22 Alur sistem.....	25
Gambar 23 Hasil proses <i>extraction</i> metode ResNext CNN	28
Gambar 24 Hasil proses klasifikasi dengan metode LSTM	28
Gambar 25 Proses pendeteksian <i>true positive</i>	36
Gambar 26 Proses pendeteksian <i>false positive</i>	37
Gambar 27 Proses pendeteksian <i>false negatif</i>	37
Gambar 28 Diagram hasil pengujian sistem.....	39

DAFTAR LAMPIRAN

Nomor	Halaman
Lampiran 1 <i>Dataset</i> video dalam bentuk <i>file .mp4</i>	47
Lampiran 2 <i>Splitting</i> video	47
Lampiran 3 <i>Face detection</i>	50
Lampiran 4 <i>Cropped face</i>	50
Lampiran 5 <i>Result</i>	54
Lampiran 6 <i>Source code</i>	56

BAB I PENDAHULUAN

1.1 Latar Belakang

Computer vision merupakan cabang ilmu yang mengarah pada pengenalan terhadap suatu objek yang dilakukan oleh komputer. Pengembangan ilmu komputer dengan konsep *computer vision* didasari dengan bagaimana cara manusia melihat objek, kemudian objek yang dilihat diterjemahkan otak untuk dikenali. Proses pendeteksian dan pelacakan menjadi area yang penting dalam *computer vision* dengan objek yang berbeda-beda (Brunett *et al.*, 2018). Implementasi dari konsep tersebut dapat dilakukan untuk deteksi objek manusia, segmentasi citra, estimasi pose manusia, klasifikasi video, dan pelacakan pergerakan objek. Keberhasilan *computer vision* akan memberikan dorongan untuk penelitian baru yang berfokus pada menemukan metode yang memiliki kinerja lebih efisien (Szegegy *et al.*, 2016).

Pendeteksi wajah adalah salah satu implementasi dari *computer vision* (Wang *et al.*, 2018). Dengan mendapatkan titik koordinat wajah pada citra akan dapat dikenali. Dengan melakukan pendeteksi wajah dapat diperoleh posisi dan ukuran wajah dalam sebuah citra, subsistem ekstraksi ciri, yang berfungsi mengekstraksi ciri-ciri yang terdapat di area wajah, dan subsistem pengenalan wajah yang bertugas membandingkan citra wajah masukan dengan sekumpulan wajah dalam suatu *database*, sehingga akhirnya dapat ditentukan tingkat pengenalan terhadap citra wajah tersebut. Permasalahan yang dapat diatasi dengan melakukan pengenalan wajah adalah mengidentifikasi pemalsuan wajah atau *deep fake* (Victoria & Solihin, 2018).

Deepfake adalah teknik untuk sintesis citra dimana seseorang dalam gambar atau video diganti dengan wajah orang lain (Pan *et al.*, 2020). Meskipun tindakan memalsukan konten bukanlah hal baru, teknologi barupun banyak yang masih sulit untuk membedakan antara asli dan palsu dalam suatu konten (Westerlund, 2019). Pemalsuan data dengan memanfaatkan *deepfake* memiliki banyak dampak salah satunya adalah munculannya data maupun berita hoaks yang tersebar di masyarakat. Data

akan membentuk persepsi, menggiring opini, dan membuat opini yang menguji kepehaman pengguna terhadap informasi yang disebar. Kemungkinan informasi yang dilahirkan kembali atau diproduksi kembali dengan tambahan opini pribadi juga menjadikan sebuah informasi yang awalnya bersifat fakta, menjadi informasi hoaks (Rahmadhany, Aldila Safitri, & Irwansyah, 2021). Pemicu terjadinya informasi hoaks dipicu dua motif yaitu ekonomi dan politik. Hoaks dapat menyebabkan perseruan hingga bukan tidak mungkin dapat menimbulkan konflik dalam kehidupan masyarakat (Najemi, Munandar, & Prayudi, 2021).

Deepfake memanfaatkan teknik *machine learning* dan kecerdasan buatan untuk memanipulasi atau menghasilkan konten visual dan audio yang digunakan untuk menyebarkan informasi palsu. Pembuatan video *deepfake* juga menggunakan *deep learning* dan metode jaringan syaraf tiruan lainnya, seperti *autoencoder* atau *Generative Adversarial Networks (GANs)*. Beberapa tahun terakhir *deepfake* menjadi permasalahan yang sering dikeluhkan oleh berbagai kalangan hingga dibutuhkan sistem untuk dapat melakukan pendeteksian *deepfake* untuk meminimalisir berita atau konten yang berbau pemalsuan dan hoaks (Suganthi *et al.*, 2022).

Penelitian mengenai *deepfake* telah dilakukan oleh berbagai peneliti, dengan menggunakan metode CNN dan *Matric Learning Function*. Dalam penelitian tersebut disebutkan bahwa metode yang digunakan akan membutuhkan data video berskala besar (Agarwal, Farid, El-Gaaly, & Lim, 2020). Selanjutnya penelitian yang mengambil objek mendeteksi pemalsuan wajah dengan menggunakan metode *Faster Region based Convolutional Neural Network (Faster R-CNN)* (Megawan, Lestari, & Halim, 2022). Ranjan dkk (Ranjan, Patil, & Kazi, 2020) mendeteksi *deepfake* pada video menggunakan CNN dan *Transfer Learning (TL)*. *Dataset* yang digunakan berasal dari video yang diekstrak menjadi *image* berukuran 128 x 128 pixel. Selanjutnya data diklasifikasi dengan dua cara yaitu menggunakan CNN, dan menggunakan CNN yang dikombinasikan dengan TL. Hasil deteksi menunjukkan bahwa akurasi tertinggi mencapai 86,49% untuk klasifikasi data yang menggunakan CNN dan TL.

Dalam penelitian ini memiliki perbedaan dari penelitian sebelumnya dari berbagai sisi. Proses identifikasi *deepfake* dilakukan dengan mengkombinasikan beberapa metode untuk proses pendeteksian *deepfake*. Proses *feature extraction* akan dilakukan dengan metode ResNext *Convolution Neural Network* dan proses klasifikasi akan dilakukan dengan metode *Long Short-Term Memory* (LSTM). Proses analisis pengujian sistem akan dilakukan dengan metode *confusion matrix*. Dengan adanya penelitian ini dengan menggunakan data *training real* dan *fake* diharapkan dapat digunakan dalam meminimalkan berbagai pemalsuan dan informasi hoaks yang beredar. Disamping itu, akan memberikan hasil analisis akurasi metode ResNext CNN dan LSTM dalam melakukan identifikasi *deepfake*.

1.2 Rumusan Masalah

Berdasarkan masalah di atas maka masalah-masalah penelitian dapat dirumuskan sebagai berikut:

1. Bagaimana cara mengolah data *image* wajah dalam video untuk mengidentifikasi *deepfake*?
2. Bagaimana kinerja dari metode *Long Short-Term Memory* dan ResNext *Convolutional Neural Networks* dalam mengidentifikasi video *deepfake*?

1.3 Tujuan Penelitian

Tujuan penelitian yang akan dicapai dalam penelitian ini adalah:

1. Untuk mengolah data *image* wajah dari video dalam mengidentifikasi video asli atau palsu.
2. Untuk mengetahui kinerja dari metode *Long Short-Term Memory* dan ResNext *Convolutional Neural Networks* dalam mengidentifikasi video *deepfake*.

1.4 Manfaat Penelitian

Penelitian ini diharapkan memberi masukan:

1. Bagi peneliti, penelitian yang diusulkan dapat menambah pengetahuan dalam mengidentifikasi video asli atau palsu berdasarkan *image* wajah.
2. Bagi institusi pendidikan magister jurusan Teknik Elektro konsentrasi Teknik Informatika, dapat digunakan sebagai referensi ilmiah dalam

mengembangkan penelitian yang berhubungan dengan identifikasi video asli atau palsu berdasarkan *image* wajah.

1.5 Batasan Penelitian

Adapun batasan-batasan dalam penelitian ini adalah;

1. Data yang digunakan dalam penelitian berasal dari video yang memiliki resolusi 360p, *frame rates* 25 fps dan berdurasi 10 detik agar hasil identifikasi *deepfake* memiliki tingkat akurasi yang tinggi dan penggunaan waktu yang efisien.
2. Metode yang digunakan adalah *Long Short-Term Memory* dan *ResNext Convolutional Neural Networks*.

1.6 Sistematika Penulisan

Adapun sistematika penulisan dari tesis ini adalah :

BAB I PENDAHULUAN

Bab I berisi penjelasan tentang latar belakang masalah yang mencakup penjelasan mengenai *computer vision*, pengenalan wajah hingga pada permasalahan pemalsuan wajah yang dapat diatasi dengan melakukan pengenalan wajah adalah mengidentifikasi pemalsuan wajah atau *deepfake*. Terdapat rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah serta sistematika penulisan yang menjadi landasan dalam penelitian ini memperoleh hasil analisis.

BAB II TINJAUAN PUSTAKA

Bab II berisi penjelasan tentang landasan teori dan penjelasan tentang hasil penelitian dari sumber acuan yang bersumber dari jurnal yang terkait dengan penelitian tentang pendeteksian *deepfake*. Dalam bab ini juga diuraikan tentang kerangka pikir tentang masalah yang sedang terjadi, dan penggunaan metode yang diusulkan sesuai dengan objek dalam penelitian yang diusulkan.

BAB III METODE PENELITIAN

Bab III ini merupakan penjelasan tentang tahapan penelitian, bagaimana pengembangan dan penerapan sistem yang diusulkan mengenai sistem untuk mengidentifikasi video berdasarkan *image* wajah *Long Short-*

Term Memory dan *ResNext Convolutional Neural Networks*, rancangan sistem dan uraian proses validasi hasil dari metode yang diusulkan.

BAB IV HASIL DAN PEMBAHASAN

Bab IV ini merupakan implementasi hasil dari proses perancangan yang telah dituliskan pada bab sebelumnya. Proses penjabaran hasil yang ditemukan pada proses implementasi akan dituliskan secara rinci pada hasil penelitian. Selanjutnya setelah hasil diperoleh akan dianalisis hingga memperoleh penjabaran dan pembahasan untuk masing-masing hasil yang ada.

BAB V KESIMPULAN DAN SARAN

Bab V merupakan bab akhir pada proses penulisan akhir penelitian. Dari proses perencanaan, implementasi, analisis pada bab sebelumnya akan diperoleh kesimpulan dan saran yang dituliskan pada bab ini.

BAB II TINJAUAN PUSTAKA

2.1 *Image Processing*

Pengolahan citra atau *image processing* adalah suatu sistem dimana proses dilakukan dengan masukan (*input*) berupa *image* dan hasilnya (*output*) juga berupa *image* (Mulyawan, Samsono, & Setiawardhana, 2011). *Image processing* salah satu jenis teknologi untuk menyelesaikan masalah mengenai pemrosesan gambar, yang saat ini juga dikembangkan menjadi *machine vision*. *Machine vision* merupakan teknologi dan metode yang dikembangkan untuk pemeriksaan dan analisis otomatis berbasis pencitraan (Ayunita, Arief, & Mirdanies, 2012). Dalam mengolah data *image* tahap yang akan dilakukan adalah sebagai berikut:

2.1.1 *Image acquisition*

Tujuan umum dari *image acquisition* adalah untuk mengubah gambar optik (data dunia nyata) menjadi sebuah *array* data numerik yang nantinya dapat dimanipulasi pada komputer, sebelum memulai pemrosesan video atau gambar, suatu gambar harus ditangkap oleh kamera dan diubah menjadi entitas yang dapat dikelola (Sindar, 2018). Proses *image acquisition* terdiri dari tiga langkah:

1. Sistem optik yang memfokuskan energi
2. Energi tercermin dari objek yang menarik
3. Sebuah sensor yang mengukur jumlah energi

2.1.2 *Pre-processing*

Pre-processing adalah suatu bentuk pengolahan atau pemrosesan sinyal dengan *input* berupa gambar (*image*) dan ditransformasikan menjadi gambar lain sebagai keluarannya dengan teknik tertentu. *Image processing* dilakukan untuk memperbaiki kesalahan data sinyal gambar yang terjadi akibat transmisi dan selama akuisisi sinyal, serta untuk meningkatkan kualitas penampakan gambar agar lebih mudah diinterpretasi oleh sistem penglihatan manusia baik dengan melakukan manipulasi dan juga penganalisisan terhadap gambar. Tujuan *pre-processing* adalah meningkatkan data gambar

yang menekan distorsi yang tidak diinginkan atau meningkatkan beberapa fitur gambar yang penting untuk diproses lebih lanjut. *Pre-processing* merupakan tahapan sebelum proses pengklasifikasian yang diperlukan untuk membersihkan, menghilangkan, mengubah sumber data menjadi data yang dapat diterima oleh sistem. Hal ini bertujuan agar data yang digunakan lebih optimal ketika digunakan pada proses pengklasifikasiannya (Muttakin & Bachtiar, 2016).

2.1.3 Feature extraction

Feature extraction adalah salah satu cara dapat digunakan untuk mengenali suatu objek berdasarkan pada histogram khusus yang dimiliki objek tersebut. *Feature extraction* bertujuan untuk melakukan perhitungan serta perbandingan yang dapat dipakai untuk klasifikasi suatu citra berdasarkan ciri-ciri histogram yang dimiliki (Religia, 2019).

2.2 Long Short Term Memory (LSTM)

Long Short Term Memory atau biasa disingkat LSTM adalah bentuk spesial dari RNN (*Recurrent Neural Network*) yang dapat melakukan pembelajaran pada dependensi jangka panjang (*long-term dependencies*). Model ini diperkenalkan oleh Hochreiter dan Schmidhuber pada tahun 1997 (Lattifia, Wira, Kadek, & Rusjayanthi, 2022). Proses metode LSTM dalam bentuk *Pseudo-Code* dapat dilihat pada Gambar 1.

Pseudo-code for the DE algorithm with LSTM classifier
 Define the size of the population NP , D dimension of problem, crossover rate c_r , scale factor F .
Initialization: Initialize the population $S_i^{t=0} = \{s_{1,i}^t, s_{2,i}^t, \dots, s_{D,i}^t\}$, $i = 1, \dots, NP$ which each individual uniformly distributed in the range $[s^{low}, s^{high}]$
 While the termination criteria is not met
 For each individual, target vector, in the population NP
 Mutation: Select three individual from the population randomly and generate a donor vector v_i^t using the following mutation equation:

$$v_{j,i}^t = s_{j,p}^t + F_i * (s_{j,r}^t + s_{j,q}^t)$$

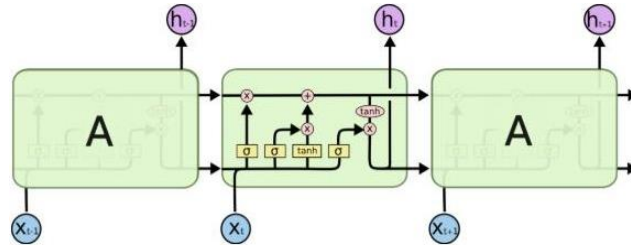
 Crossover: Compute the trial vector for the i th target vector $u_{j,i}^{t+1}$:

$$u_{j,i}^t = \begin{cases} v_{j,i}^t & \text{if } r_i \leq c_r \text{ or } j = J_{rand} \\ s_{j,i}^t & \text{otherwise} \end{cases}$$

 Selection: Apply LSTM classifier as fitness function f and evaluate s_i^t and u_i^t :
 If $f(s_i^t) \leq f(u_i^t)$ then $s_i^{t+1} = u_i^t$
 Else $s_i^{t+1} = s_i^t$
 End For
 End While

Gambar 1 *Pseudo-Code* klasifikasi metode LSTM (Nakisa, Rastgoo, Rakotonirainy, Maire, & Chandran, 2018)

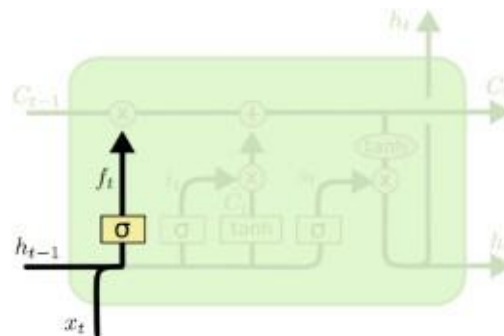
Semua *Recurrent Neural Network* memiliki bentuk rangkaian modul jaringan syaraf yang berulang. LSTM juga memiliki struktur yang sama namun memiliki tambahan fitur berupa gerbang pada sel (Wiranda & Sadikin, 2019).



Gambar 2 Detail struktur LSTM

LSTM akan menentukan informasi apa yang akan dibuang dari sel. Keputusan ini dibuat oleh *forget gate layer*. Layer ini akan memperhatikan h_{t-1} dan x_t sehingga akan menghasilkan keluaran antara 0 dan 1. Keluaran 0 merepresentasikan bahwa informasi akan dilupakan sedangkan keluaran 1 merepresentasikan bahwa informasi tidak akan dilupakan.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

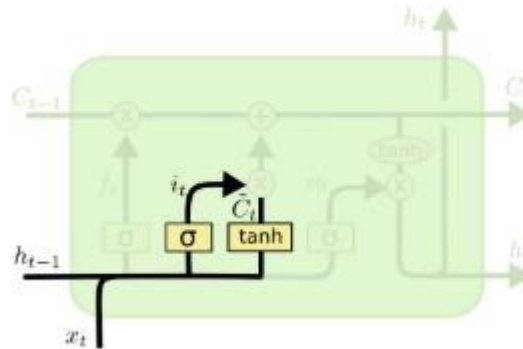


Gambar 3 Struktur *forget layer*

Langkah selanjutnya adalah menentukan apakah informasi akan disimpan pada sel. Pertama, sebuah *layer* sigmoid bernama “*input gate layer*” menentukan nilai mana yang akan diperbarui. Selanjutnya, sebuah *layer* tanh membuat vektor dari nilai kandidat baru, C_t , yang dapat ditambah ke *state*. Langkah selanjutnya, kedua *layer* ini akan dikombinasikan untuk memperbarui *state*.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

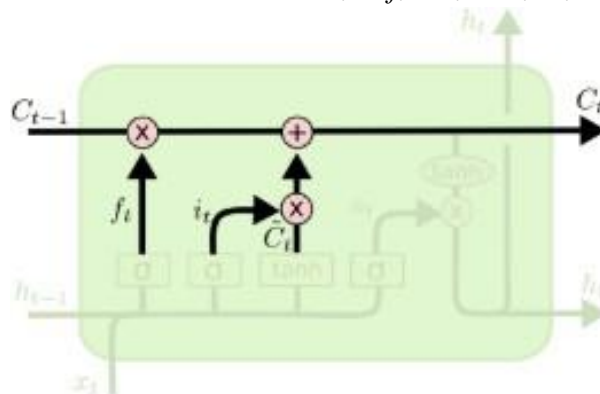
$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$



Gambar 4 Struktur *remember gate*

Selanjutnya, *state* lama akan diperbarui, C_{t-1} ke *state* sel baru C_t . Kemudian, f_t akan dikalikan dengan *state* lama dengan mengabaikan informasi yang sudah dilupakan sebelumnya. Lalu, i_t ditambahkan dengan C_t .

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

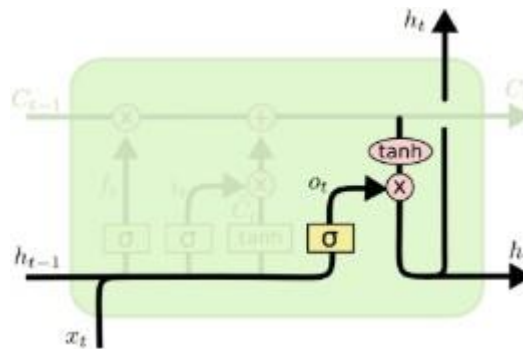


Gambar 5 Struktur *update layer*

Langkah terakhir adalah menentukan apa keluarannya. Pertama, *layer* sigmoid akan menentukan bagian dari sel yang akan dikeluarkan. Kemudian, sel tersebut akan dilewatkan pada *layer* tanh (untuk memaksa nilai keluaran diantara -1 dan 1) dan mengalikan dengan keluaran dari gerbang sigmoid.

$$O_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (5)$$

$$H_t = o_t * \tanh(C_t) \quad (6)$$

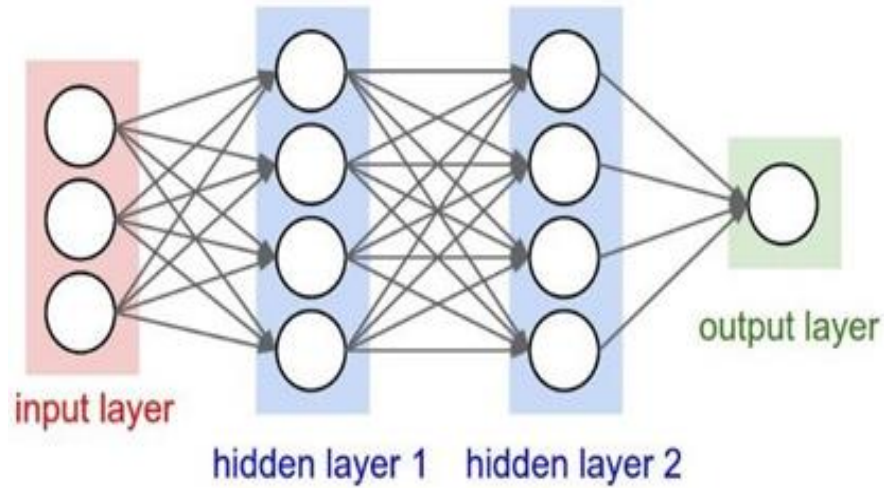


Gambar 6 Struktur *output layer*

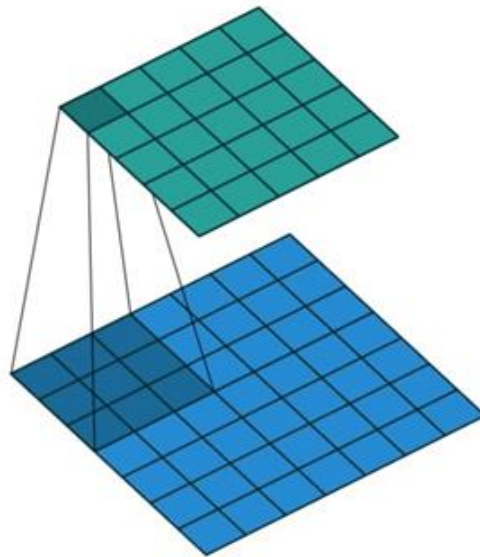
2.3 ResNext Convolutional Neural Network (ResNext CNN)

Convolutional Neural Network (CNN) merupakan pengembangan dari *Multilayer Perceptron* (MLP) yang dirancang untuk mengolah data dua dimensi. CNN termasuk dalam jenis *Deep Neural Network* karena kedalaman jaringannya yang tinggi dan banyak diterapkan pada data citra. Pada umumnya *convolutional layer* akan mendeteksi fitur berupa *edge* pada citra, kemudian *subsampling layer* akan memperkecil dimensi fitur yang didapat dari *convolutional layer*, dan terakhir meneruskannya ke *output node* melalui proses *forward propagation*. MLP kurang sesuai untuk digunakan karena tidak menyimpan informasi spasial dari data citra dan menganggap setiap piksel adalah fitur yang independen sehingga menghasilkan hasil yang kurang baik. CNN pertama kali dikembangkan dengan nama *NeoCognitron* oleh Kunihiko Fukushima, seorang peneliti dari *NHK Broadcasting Science Research Laboratories*, Kinuta, Setagaya, Tokyo, Jepang. Konsep tersebut kemudian dimatangkan oleh Yann LeChun, seorang peneliti dari *AT&T Bell Laboratories* di Holmdel, New Jersey, USA. Model CNN dengan nama *LeNet* berhasil diterapkan oleh LeChun pada penelitiannya mengenai pengenalan angka dan tulisan tangan. Pada tahun 2012, Alex Krizhevsky dengan penerapan CNN miliknya berhasil menjuarai kompetisi *ImageNet Large Scale Visual Recognition Challenge 2012*. Prestasi tersebut menjadi momen pembuktian bahwa metode *deep learning*, khususnya CNN. Metode CNN terbukti berhasil mengungguli metode *machine learning* lainnya seperti SVM pada kasus klasifikasi objek pada citra (Anjani, Pratiwi, & Norfa Bagas Nurhuda, 2021) (Zhang et al., 2019) (Saleem et al., 2022).

Cara kerja CNN memiliki kesamaan pada MLP, namun dalam CNN setiap neuron dipresentasikan dalam bentuk dua dimensi, tidak seperti MLP yang setiap neuron hanya berukuran satu dimensi.

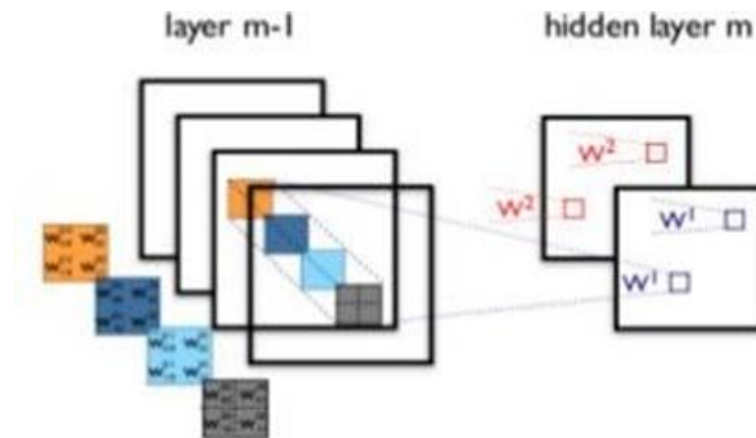


Gambar 7 Arsitektur MLP sederhana



Gambar 8 Ilustrasi proses konvolusi

Karena sifat proses konvolusi, maka CNN hanya dapat digunakan pada data yang memiliki struktur dua dimensi seperti citra dan suara.



Gambar 9 Proses konvolusi pada ResNext CNN

ResNext CNN adalah salah satu pengoptimalan dari CNN dalam mendapatkan hasil yang lebih optimal. ResNext CNN salah satu metode yang memiliki keunggulan dalam mempertahankan kompleksitas komputasi. Memiliki kinerja yang baik dalam mengurangi 6,7% kesalahan dalam melakukan pemrosesan data (Annapurani & Ravilla, 2019). Metode ini memiliki desain yang sederhana dalam melakukan *extraction* (Xie, Girshick, Dollár, Tu, & He, 2017) *et al.* (Pant, Yadav, & Gaur, 2020). Dalam penelitian ini metode ResNext CNN digunakan dalam proses ekstraksi fitur, hasil dari ekstraksi akan diklasifikasi menggunakan metode LSTM.

2.4 Evaluasi Proses Pendeteksian

Proses pengujian klasifikasi dapat dilakukan dengan menggunakan metode *confusion matrix*. *Confusion matrix* adalah metode yang menggambarkan performa dari sebuah model atau metode klasifikasi secara spesifik. Dalam tabel *confusion matrix* memiliki baris dan kolom yang memiliki arti masing-masing. Setiap baris dari *matrix* yang ada akan menggambarkan kelas aktual dari data, dan setiap kolom menggambarkan kelas prediksi dari data.

Confusion matrix berisi semua informasi mentah tentang prediksi yang dilakukan oleh sebuah model klasifikasi pada kumpulan data tertentu. Untuk mengevaluasi akurasi generalisasi model, adalah umum untuk menggunakan kumpulan data pengujian yang tidak digunakan selama proses pembelajaran model tersebut (Caelen, 2017). *Matrix* dari proses pengujian data dilihat pada Tabel 1 (Li, Li, Li, Yao, & Xiao, 2021):

Tabel 1 *Confusion matrix*

	<i>Prediction Condition</i>		
<i>Actual Condition</i>	<i>Total population</i> = $P + N$	<i>Positive (P)</i>	<i>Negative (N)</i>
	<i>Positive (P)</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	<i>Negative (N)</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

True Positive (TP) = Jumlah banyak data yang aktual kelasnya positif dan memprediksi positif.

True Negative (TN) = Jumlah banyak data yang aktual kelasnya negatif dan memprediksi negatif.

False Positive (FP) = Jumlah banyak data yang aktual kelasnya positif, namun hasil klasifikasi negatif.

False Negative (FN) = Jumlah banyak data yang aktual kelasnya negatif, namun hasil klasifikasi positif.

Pengujian akan dilakukan dengan melihat nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dari keseluruhan data percobaan. *Accuracy* adalah jumlah keseluruhan seberapa sering metode melakukan klasifikasi secara benar. *Precision* adalah jumlah dari metode dalam melakukan klasifikasi secara benar, seberapa sering klasifikasi itu benar. *Recall (Sensitivity/True Positive Rate)* proses ketika aktualnya positif, seberapa sering metode melakukan klasifikasi positif. *F1-Score* adalah nilai rata-rata harmonik dari *Precision* dan *Recall*. Formula dari pengujian pada nilai *Precision*, *Recall*, *Accuracy* dan *F1-Score* dapat dilihat pada Persamaan 7,8,9,10 (Saputro & Sari, 2020) (Naiemi, Ghods, & Khalesi, 2019):

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

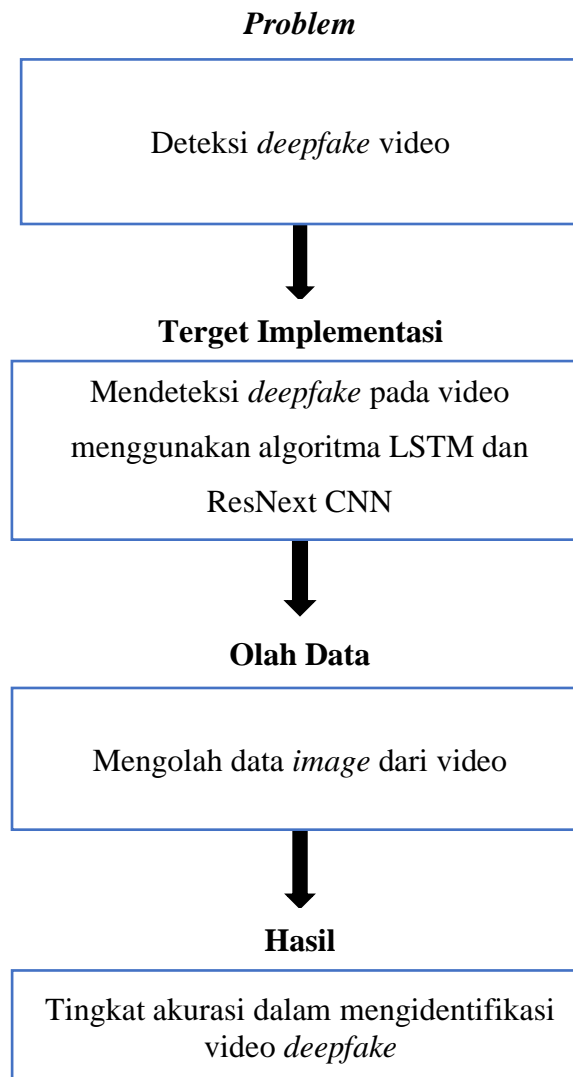
$$F1-Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (10)$$

2.5 State of the Art

Tabel 2 *State of the Art*

No	Penulis / Tahun	Judul	Metode	Hasil
1	Tech, Seth dan Tech / 2021	<i>Detection of Deep-fakes in Videos using CNN and Transformers</i>	Gabungan CNN dan Transformer (CViT)	Kombinasi dari CNN dan model <i>trafo Vision</i> digunakan untuk mengklasifikasikan jika video menjadi sasaran manipulasi. Dan memvalidasi model kami pada kumpulan data besar video palsu dan asli yang dihasilkan CViT dengan kurasi = 91.25%
2	Guera dan Delp / 2019	<i>Deepfake Video Detection Using Recurrent Neural Networks</i>	RNN	Klasifikasi menunjukkan bahwa klasifikasi data dengan secara akurat apakah video telah dimanipulasi atau tidak dengan data video sesedikit 2 detik
3	Irene Amerini dan Roberto Caldelli / 2020	<i>Exploiting Prediction Error Inconsistencies through LSTM-based Classifiers to Detect Deepfake Videos</i>	LSTM-based Classifiers	LSTM yang memiliki kemampuan untuk mengeksploitasi korelasi temporal seperti itu telah menunjukkan kinerja yang menjanjikan. Hasil yang didapat memuaskan walaupun untuk yang sangat video terkompresi penurunan akurasi terdaftar. Masa depan pengembangan akan dikhususkan untuk mengoptimalkan metode kami untuk terintegrasi dalam hubungannya dengan berbasis teknologi

2.6 Kerangka Pikir



Gambar 10 Kerangka pikir

Pada Gambar 10, menjelaskan alur penelitian yang akan dilakukan. Pada tahap pertama menjelaskan permasalahan yang ada yaitu munculnya berita atau informasi hoaks yang beredar dalam masyarakat dalam bentuk video yang telah diubah keasliannya. Solusi yang akan ditawarkan adalah sebuah sistem yang mampu mendeteksi *deepfake* pada video dengan menggunakan metode ResNext CNN dan LSTM yang diharapkan nantinya dapat mendeteksi keasliannya berdasarkan wajah pada video. Pengujian pada sistem dilakukan dengan cara menguji sistem yang telah dibuat untuk melihat tingkat keakuratan sistem.