



**Deteksi Video *Deepfake*
berdasarkan ketidakcocokan area
mulut berbasis objek**

Oleh:

Rayhan Khalif Putra (NIM: 1301170604)

Latar Belakang

- Perhatikan kedua video ini, Video kedua merupakan video yang dimanipulasi dengan *Deepfake*



Studi Terkait

Shurti.Agarwal,hfarid.

Ohad,Maneesh

*Detecting “Deepfake Videos From
Phoneme-Viseme Mismatches”*

2020

- **Kelebihan dan kekurangan**

Akurasi penelitian sudah cukup tinggi .Tetapi, Penelitian pada jurnal ini masih sangat terbatas. Viseme yang digunakan masih hanya M,B,P. Dan dataset hanya menggunakan 1 figure saja

- **Deskripsi**

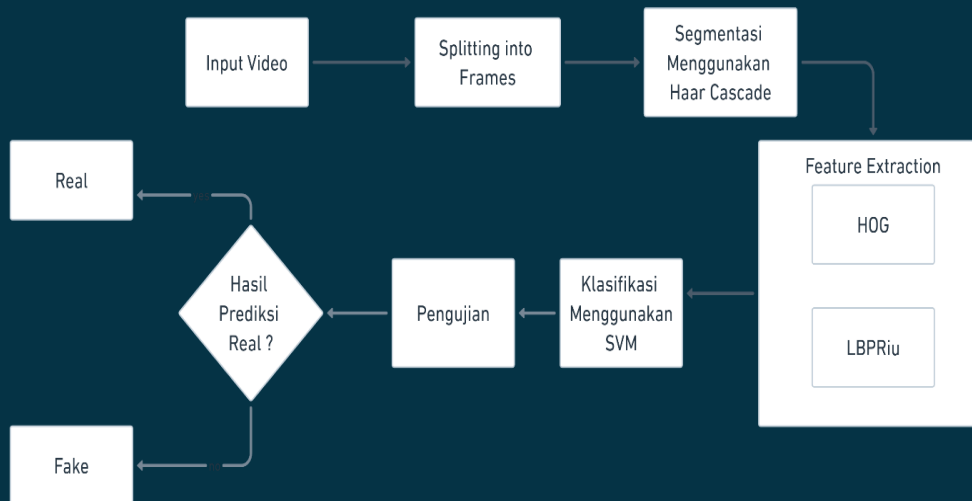
Jurnal ini melakukan penelitian untuk mendeteksi video deepfake berdasarkan ketidakcocokan viseme M,B,P menggunakan 2 metode yaitu Profiling dan CNN

- **Permasalahan**

Akurasi deteksi video *Deepfake* berdasarkan ketidakcocokan area mulut masih bisa ditingkatkan. Dan metode yang digunakan bisa lebih dieksplor dari referensi jurnal sebelumnya.

Sistem yang Diusulkan

- Membangun sistem untuk mendeteksi video *Deepfake* dengan fokus pada ketidakcocokan area mulut dengan kombinasi metode ekstraksi ciri :
 1. Local Binary Pattern Rotation Invariant Uniform
 2. Histogram of Oriented Gradient
- Metode Klasifikasi menggunakan *Support Vector Machine (SVM)*



Model yang Dibangun

1. Input Video dan Splitting menjadi Frame

- Melakukan Input Video
- Pemisahan video menjadi frame

$$N = T \times F$$

N = Jumlah total frame dalam video.

T = Durasi video dalam detik.

F = Frame rate video, yaitu jumlah frame yang diambil per detik (fps).

$$t_i = \frac{i}{F}$$

t_i : Waktu di mana frame ke- i diambil, dalam detik.

i : Indeks dari frame yang dimulai dari 0 hingga $N - 1$.



Model yang Dibangun

2. Segmentasi Menggunakan *Haarcascade*

- Segmentasi *Haarcascade* terdiri dari 4 tahap :
 - A. A.Fitur Haarlike
 - B. Integral Image
 - C. Adaboost
 - D. Cascade Classifier

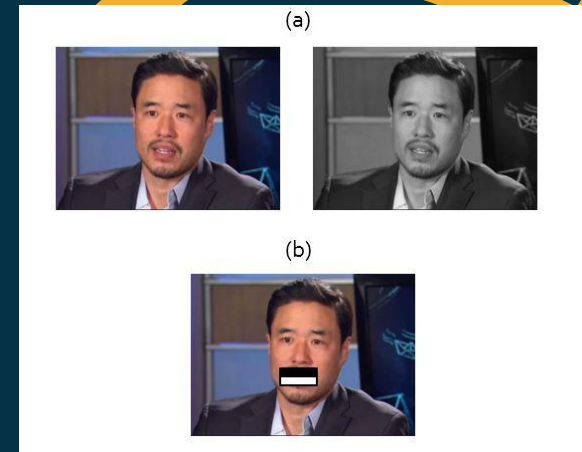
A. Fitur Haarlike

1. Citra Greyscale

$$I_{gray}(x, y) = 0.299 \times I_R(x, y) + 0.587 \times I_G(x, y) + 0.114 \times I_B(x, y)$$

2. Deteksi tepi area gambar

$$H = \sum_{(x,y) \in R_w} I(x, y) - \sum_{(x,y) \in R_b} I(x, y)$$



Model yang Dibangun

2. Segmentasi Menggunakan *Haarcascade*

B. Integral Image

- Transformasi citra matriks menjadi citra integral

$$L(x, y) = \sum_{i=0, j=0}^{i \leq x, j \leq y} I(i, j), 0 \leq x \leq N, 0 \leq y \leq M,$$

- Setiap Elemen matriks dihitung menggunakan relasi rekursif

$$L(x, y) = I(x, y) - L(x - 1, y - 1) + L(x, y - 1) + L(x - 1, y)$$

C. Adaboost

- Proses filtrasi sebagai pencocokan data masukan citra dengan kumpulan data training

D. Cascade Classifier

- Proses klasifikasi untuk mencari citra wajah dan mulut berdasarkan hasil data latih dari adaboost

Model yang Dibangun

3. Fitur Ekstraksi Local Binary Pattern Rotation Invariant Uniform

- Mengubah citra RGB menjadi Grayscale
- Membagi citra menjadi beberapa blok
- Operator Circular menggunakan rasio (8,1)
- Menghitung LBPriu2 dengan formulasi :

$$LBP_{P,R}^{riu2} = \int \sum_{p=0}^{p-1} s(g_p - g_c) \quad \begin{array}{l} \text{jika } U(LBP_{P,R}) \leq 2 \\ \text{kebalikannya} \end{array}$$



Model yang Dibangun

4. Fitur Ekstraksi Histogram of Oriented Gradient

- Menghitung nilai gradien citra
- Menghitung koordinat sumbu
- Menghitung histogram dari orientasi tiap sel
- Nilai HOG dinormalisasi

$$|G| = \sqrt{I_x^2 + I_y^2}$$

$$\theta = \arctan\left(\frac{I_x}{I_y}\right)$$

$$b = \frac{b}{\sqrt{b^2 + c}}$$

$$w = \frac{w}{\sqrt{\|w\|^2 + c}}$$



Model yang Dibangun

5. Klasifikasi Support Vector Machine

- Memisahkan dua kelas dengan menemukan *Hyperlane* terbaik
- Fitur yang digunakan dari hasil fitur ekstraksi adalah 1773
- Bobot (w) akan memiliki 1773 nilai
- Membangun model SVM

$$\frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2 + w_3^2 + \dots + w_{1773}^2)$$

$$y_i(w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + \dots + w_{1773} \cdot x_{1773}) \geq 1 \quad \text{untuk kelas } real$$

$$y_i(w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + \dots + w_{1773} \cdot x_{1773}) \leq -1 \quad \text{untuk kelas } fake$$

$$(w_1 \cdot x_1)(w_2 \cdot x_2)(w_3 \cdot x_3) \dots (w_{1773} \cdot x_{1773}) + b \geq 1$$

$$(-w_1 \cdot x_1)(-w_2 \cdot x_2)(-w_3 \cdot x_3) \dots (-w_{1773} \cdot x_{1773}) - b \geq 1$$

$y = +1$ atau *REAL*, jika $w \cdot z + b > 0$

$y = -1$ atau *FAKE*, jika $w \cdot z + b < 0$

DataSet

Terdapat Dataset yang digunakan berasal dari dataset yang sudah ada. Dataset yang digunakan yaitu Celeb-DF versi 2 (Celeb-DF v2). Dataset ini terdiri dari video *REAL* dan deep*FAKE* yang memiliki kualitas visual yang setara dengan video yang berderdar secara online yang di *convert* menjadi frame image. Dataset ini terdiri dari data *REAL* dan data *FAKE* :

Jenis Gambar	Jumlah Gambar
FAKE	1339
REAL	1339
Total	2678



Skenario Pengujian

- Skenario pengujian terdiri dari 4 video yang menjadi data uji pada penelitian ini
 1. Video real pria *Interview*
 2. Video fake pria *Interview*
 3. Video real wanita *Interview*
 4. Video fake wanita *Interview*



METRIK EVALUASI

- Precision : Mengukur seberapa baik model membuat prediksi yang benar untuk kelas positif dari total prediksi positif

$$Precision = \frac{TP}{TP+FP}$$

- Recall : Mengukur seberapa baik model dalam mengidentifikasi kelas positif dengan benar

$$Recall = \frac{TP}{TP+FN}$$



METRIK EVALUASI

- F1-Score : Menjelaskan keseimbangan antara precision dan recall.

$$F1-Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

- Akurasi : Mengukur kesesuaian hasil pengukuran dengan standar atau nilai yang diterima sebagai benar

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$



METRIK EVALUASI

- TP = sampel video yang benar terklasifikasi sebagai video deepfake
- TN = sampel video yang benar terklasifikasi sebagai video asli
- FP = sampel video yang salah terklasifikasi sebagai video deepfake (padahal asli)
- FN = sampel video yang salah terklasifikasi sebagai video asli (padahal deepfake)



Evaluasi Performansi

- Nilai akurasi dari model klasifikasi sebesar 99.79 % dan dengan nilai precision, recall, f1-score.

Kategori	Precision	Recall	F1-score	Support
FAKE	1.00	1.00	1.00	284
REAL	0.99	1.00	1.00	192
Akurasi				99.79%
Macro Avg	1.00	1.00	1.00	476
Weighted Avg	1.00	1.00	1.00	476



KESIMPULAN

Berdasarkan hasil evaluasi dari berbagai video dan kategori, ditemukan bahwa model-model yang diuji menunjukkan performa yang baik dalam klasifikasi. Akurasi model yang bervariasi antara 66.67% hingga 100%. Secara keseluruhan, rata-rata akurasi model mencapai 90.21%, mengindikasikan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan video dengan benar. Tetapi untuk kasus *Deepfake* akurasi yang dihitung memang harus terindikasi 100% agar bisa dikatakan valid apakah video tersebut merupakan video asli atau *Deepfake*.

Referensi

- [1] Shurti.Agarwal,hfarid. Ohad,Maneesh, “Detecting Deepfake Videos From Phoneme-Viseme Mismatch,” *IEEE International Conference on Computer Vision*.2020
- [2] Azwar, “ Integrasi Ekstraksi Fitur Local Binary Pattern dan Gray-Level Coocurence Metrix untuk Pengenalan Eksopresi Mulut Pembelajar.” ISSN online 2548-7779. 2017
- [3] Zhao.Guoying, “Rotation-Invariant Image and Video Description With Local Binary Pattern Features.”*IEEE Transaction on Image Processing*, Vol.21,No.4 2012
- [4] Viola.Paul, Jones.Michael J, “Robust Real-Time Face Detection” *International Journal of Computer Vision* 57(2).137-154. 2004
- [5] N.Dalal, B.Triggs. “Histogram of Oriented Gradients for Human Detection.” *IEEE Computer Vision and Pattern Recognition*, Volume 1, Pages 886-893 2005



Referensi

[6] K.Durgesh, B.Lekha. “ Data Classification Using Support Vector Machine” Journal of Theoretical and Applied Information Technology. 2010





Thank You!