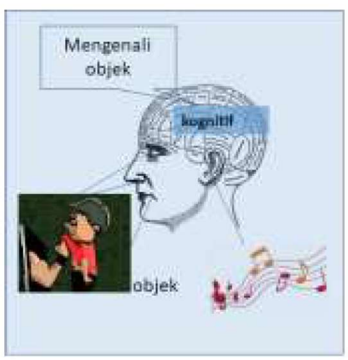
DETEKSI TINDAK KEKERASAN MENGGUNAKAN DEEP TRANSFER LEARNING

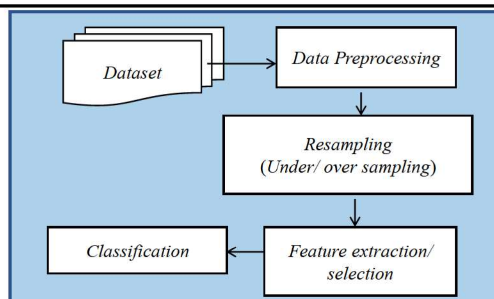
Dr. Elly Matul Imah

Universitas Negeri surabaya

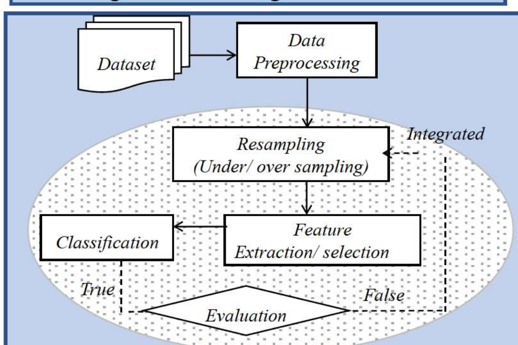
1

Pengenalan Pola





Blok diagram sistem Pengenalan Pola Klasik



Blok diagram sistem Pengenalan Pola Saat Ini

Pengenalan pola yang dilakukan manusia merupakan proses pembentukan persepsi dari suatu objek atau kejadian kemudian mengolahnya menjadi sesuatu yang memiliki arti secara keseluruhan.

Kemampuan manusia untuk mengenali pola bergantung pada naluri dan sistem kognitif yang dimiliki.

Dalam Ilmu Komputer khususnya bidang pembelajaran mesin, pengenalan pola adalah proses pembelajaran pada mesin untuk mengenali atau mengklasifikasikan objek berdasarkan data pelatihan

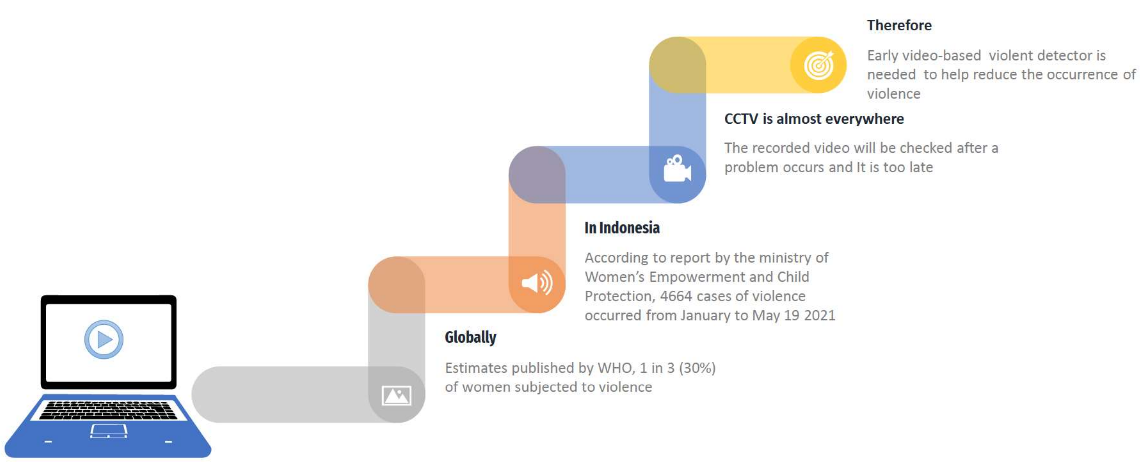
2

"Penggunaan kekuatan fisik atau kekuasaan dengan sengaja, ancaman atau tindakan nyata, terhadap diri sendiri, orang lain, atau terhadap kelompok atau komunitas, yang mengakibatkan atau memiliki kemungkinan besar mengakibatkan cedera, kematian, kerugian psikologis, perkembangan yang salah atau perampasan."

—World Health Organization

3

Introduction cont.



4

Pada Perempuan:
Secara global sekitar 1 dari 3 (30%) wanita di seluruh dunia telah mengalami kekerasan fisik dan/atau seksual pasangan intim atau kekerasan seksual non-pasangan dalam hidup mereka.

Pada Remaja: Kekerasan remaja adalah masalah kesehatan masyarakat global. Ini mencakup berbagai tindakan mulai dari intimidasi dan perkelahian fisik, hingga serangan seksual dan fisik yang lebih parah hingga pembunuhan

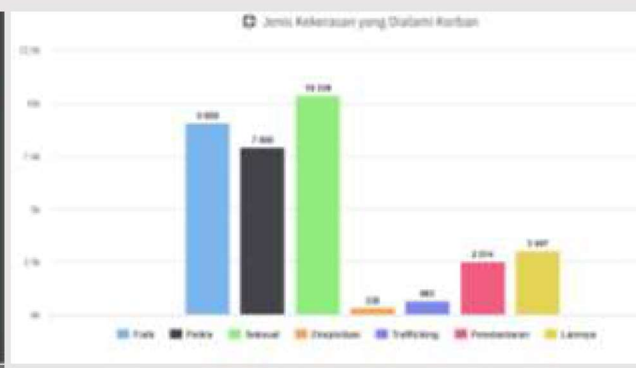
Pada Anak:
Setidaknya 1 miliar anak berusia 2–17 tahun, telah mengalami kekerasan atau penelantaran fisik, seksual, atau emosional dalam satu tahun terakhir

Pada Lansia: Sekitar 1 dari 6 orang berusia 60 tahun ke atas mengalami beberapa bentuk pelecehan di lingkungan masyarakat selama setahun terakhir.

5

KASUS TINDAK KEKERASAN DI INDONESIA

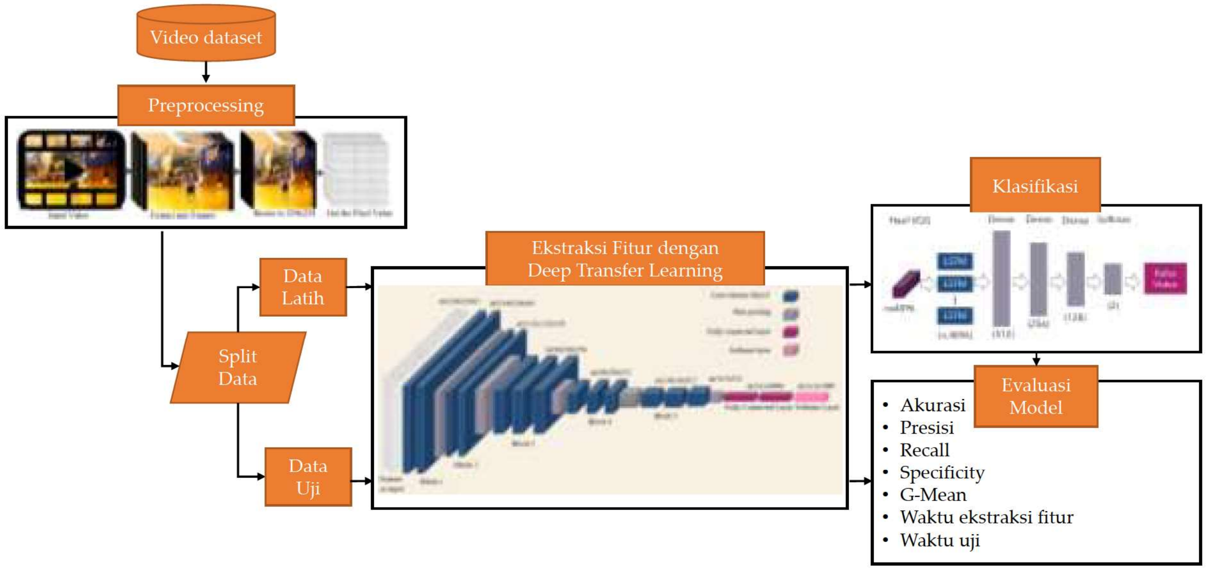
Grafik berdasarkan data dari Badan Peradilan Agama dan data formulir angket yang diterima Komnas Perempuan dari tahun ke tahun. Sumber: CATAHU2021: Catatan Tahunan Kekerasan Terhadap Perempuan Tahun 2020 Komnas Perempuan



Grafik berdasarkan data pada https://kekerasan.kemenpppa.go.id/ di tahun 2021 dengan jumlah kasus 25228. Pada tahun 2022 hingga tanggal 19 Januari 2022, sudah terdapat 1428 kasus

6

Methode



7

TRANSFER LEARNING

Transfer learning is an approach in deep learning (and machine learning) where knowledge is transferred from one model to another.

Def:
Model A is successfully trained to solve source task T.a using a large dataset D.a.
However, the dataset D.b for a target task T.b is too small, preventing Model B from training efficiently.
Thus, we use part of model A to predict results for task T.b.
(Andrew Ng)

A common misconception is that training and testing data should come from the same source or be with the same distribution.

8

Deep Transfer Learning

What is Deep Transfer Learning?
Generally, refers to a process where a model trained on one problem is used in some way on a second related problem

Visual Geometry Group Network-16
VGG16 is a convolution neural net (CNN) architecture which was used to win ILSVR(Imagenet) competition in 2014.

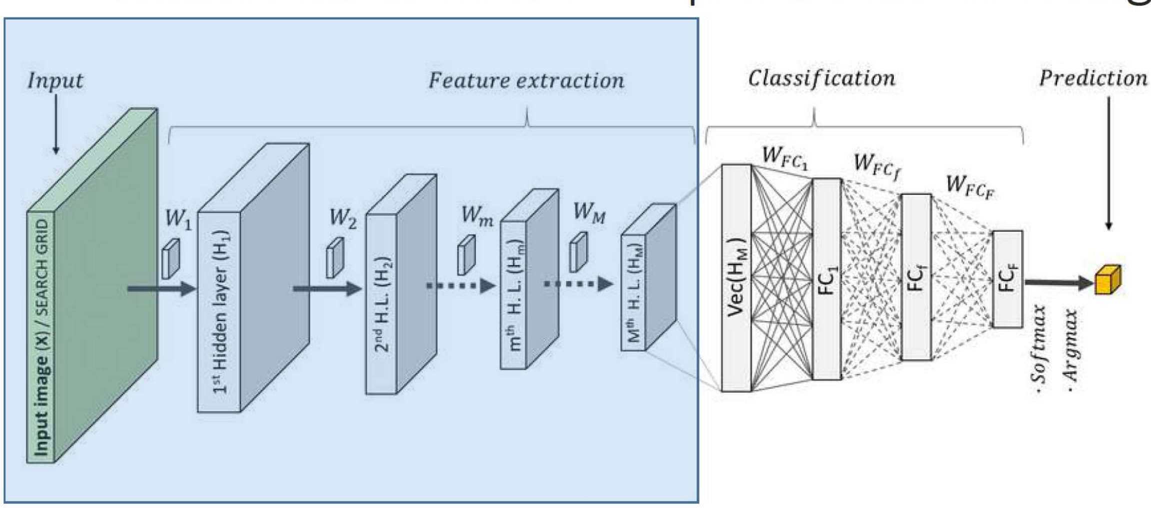
Excellent vision model architecture
The model achieves 92.7% top-5 test accuracy in ImageNet

Handle huge data sets
Because it contains 16 weighted layer with 138357544 parameters.

unique thing about VGG16
Focused on having convolution layers of 3x3 filter with a stride 1 and always used same padding and maxpool layer of 2x2 filter of stride 2.

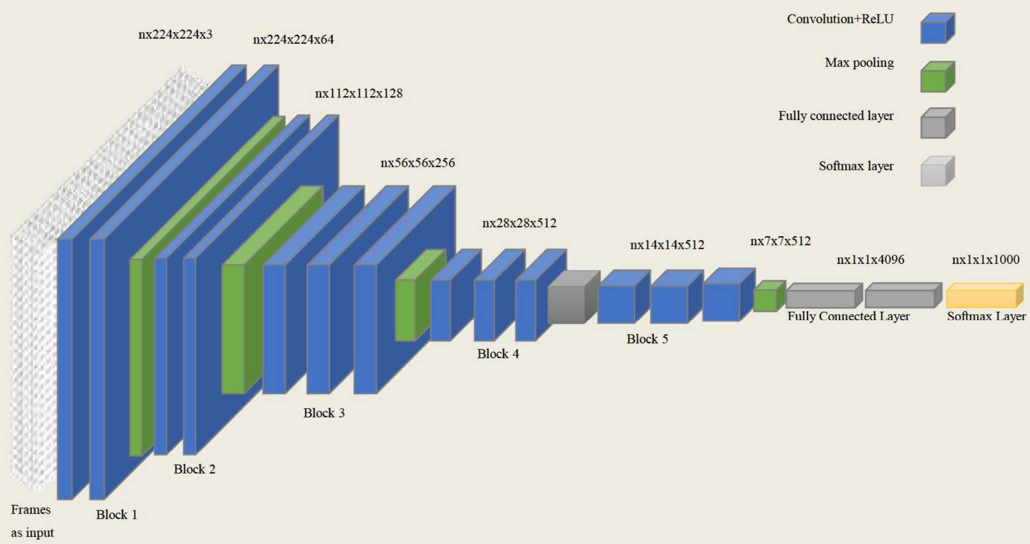
9

Gambaran mudah dari Deep Transfer Learning



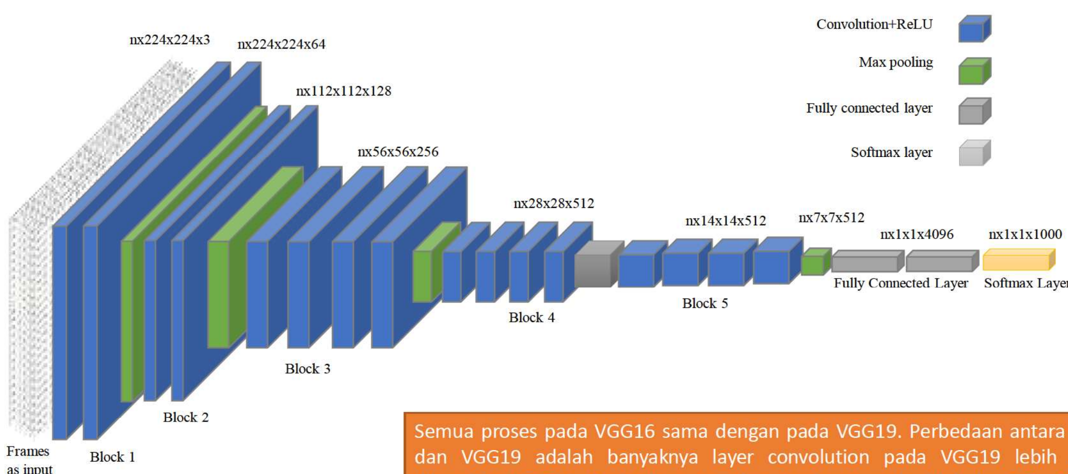
10

VISUAL GEOMETRY GROUP-16



11

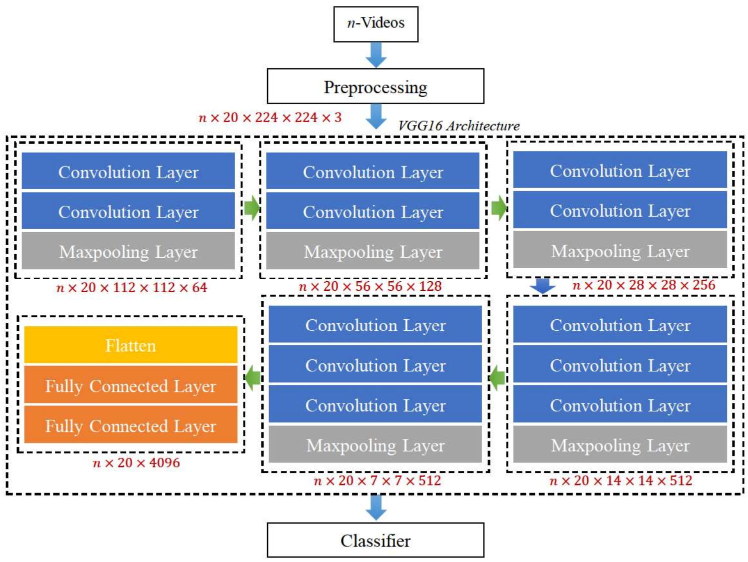
Visual geometry group-19



Semua proses pada VGG16 sama dengan pada VGG19. Perbedaan antara VGG16 dan VGG19 adalah banyaknya layer convolution pada VGG19 lebih banyak dibanding VGG16. Perbedaan tersebut terdapat pada blok ke 3-5. Sehingga parameter yang digunakan pada VGG19 lebih banyak.

12

VGG16 Feature Extraction

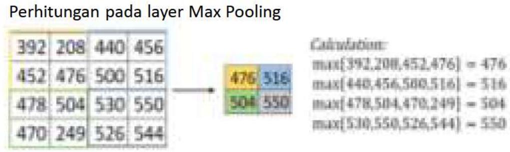


13

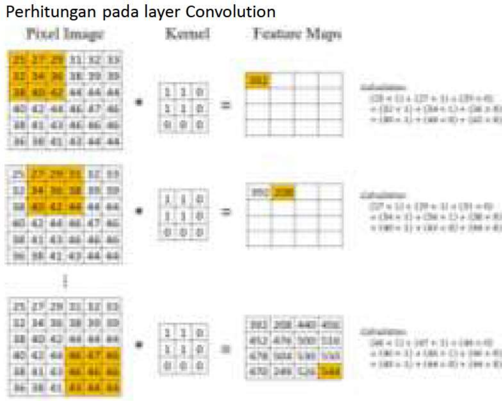
Gambaran Perhitungan pada VGG-16

Input berupa nilai pixel dari n gambar berwarna

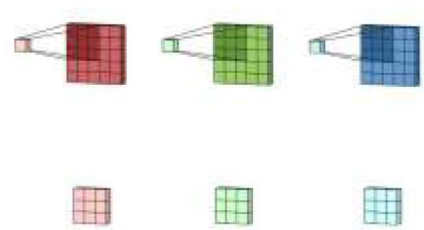
Perhitungan pada layer Max Pooling



Perhitungan pada layer Convolution



Ilustrasi konvolusi pada gambar RGB



Source: https://towardsdatascience.com/a-comprehensive-introduction-to-different-types-of-convolutions-in-deep-learning-669281e58215#:~:text=A%20%E2%80%9CKernel%E2%80%9D%20refers%20to%20a,is%20a%20collection%20of%20kernels.

14

DETIL VGG16

Data input berupa nilai pixel dari n gambar dalam format RGB berukuran 224x224.

Setelah melalui proses konvolusi dan max pooling pada blok pertama dihasilkan data berdimensi nx112x112x64

Proses konvolusi dan max pooling dilakukan kembali pada blok ke-2 dan diperoleh data berdimensi nx28x28x512

Pada blok ke-3 dihasilkan data berdimensi 28x28x256.

Pada blok ke-4 dan ke-5 dihasilkan data berdimensi 14x14x512 dan 7x7x512

Kemudian data melalui layer flatten. Data diubah menjadi berdimensi nx25088

Selanjutnya data melalui dua fully connected layer dan diperoleh data berdimensi nx4096

Pada awalnya VGG digunakan untuk mengklasifikasikan data dengan 1000 kelas oleh karena itu layer softmax memiliki 1000 unit. Layer softmax ini tidak kami gunakan, sebagai gantinya data yang telah diperoleh sebesar nx4096 digunakan sebagai input pada klasifier yang digunakan

Description	Layers	Output shape	Parameter
Input	Input	(None, 224, 224, 3)	0
Block 1	2D Convolution	(None, 224, 224, 64)	1792
	2D Convolution	(None, 224, 224, 64)	36928
	2D Max Pooling	(None, 112, 112, 64)	0
Block 2	2D Convolution	(None, 112, 112, 128)	73856
	2D Convolution	(None, 112, 112, 128)	147584
	2D Max Pooling	(None, 56, 56, 128)	0
Block 3	2D Convolution	(None, 56, 56, 256)	295168
	2D Convolution	(None, 56, 56, 256)	590080
	2D Convolution	(None, 56, 56, 256)	590080
	2D Max Pooling	(None, 28, 28, 256)	0
Block 4	2D Convolution	(None, 28, 28, 512)	1180160
	2D Convolution	(None, 28, 28, 512)	2359808
	2D Convolution	(None, 28, 28, 512)	2359808
	2D Max Pooling	(None, 14, 14, 512)	0
Block 5	2D Convolution	(None, 14, 14, 512)	2359808
	2D Convolution	(None, 14, 14, 512)	2359808
	2D Convolution	(None, 14, 14, 512)	2359808
	2D Max Pooling	(None, 7, 7, 512)	0
	Flatten	(None, 25088)	0
	Fully Connected	(None, 4096)	102764544
	Fully Connected	(None, 4096)	16781312
Predictions	Softmax	(None, 1000)	4097000
Total Parameters			138357544

15

15

Support Vector Machine

Kernels:

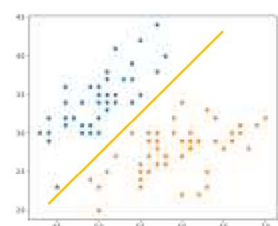
Linear $K(x,z) = xz$

RBF $k(x,z) = exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$

Polynomial $K(x,z) = (x^T z + c)^3$

Outstanding performance
solving data classification problems in the small sample, nonlinear, and high dimensional feature space

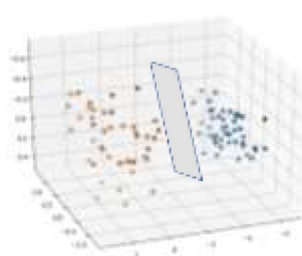
Core idea
find a separating hyper-plane that will correctly distinguish the data



Hyperplane in $\mathbb{R}^2$ is a line

Regularization parameter $C = 1.0$
determine the tradeoff of model complexity and classification error

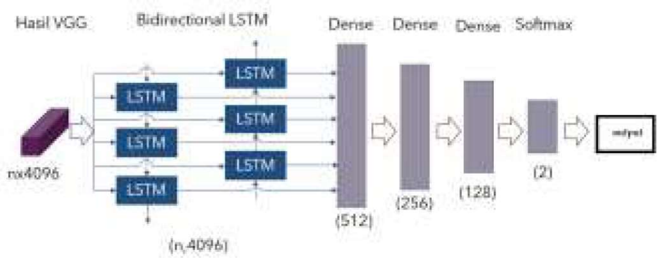
$\gamma = 1/(n\ features \times var\ data)$
the kernel coefficient for RBF and polynomial

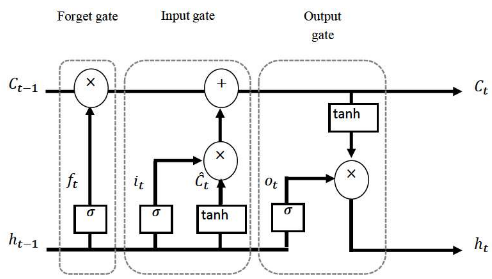


Hyperplane in $\mathbb{R}^3$ is a plane

16

Long-Short Term Memory



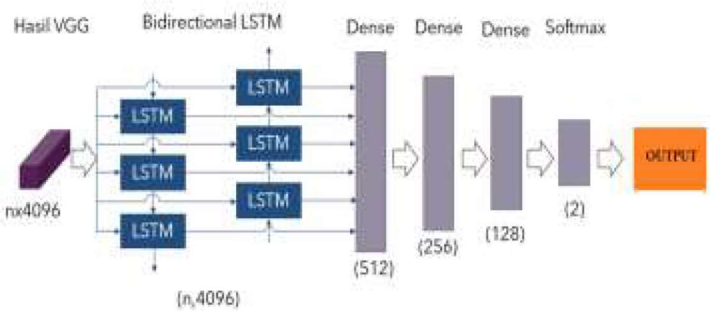


- Memiliki 3 gate: input gate, forget gate and output gate
- Mampu memodelkan dependensi urutan jangka Panjang

- Forget gate: Menentukan informasi apa yang akan disimpan dan dilupakan.
  $$f_t = \sigma(x_t U^f + h_{t-1} W^f)$$
- Input gate: memperbarui konten atau isi sel memori.
  $$i_t = \sigma(x_t U^i + h_{t-1} W^i)$$
  $$\hat{C}_t = \tanh(x_t U^g + h_{t-1} W^g)$$
- Pada tahap berikutnya, status lama C_{t-1} diaktualisasi dengan C_t dengan cara mengalikannya status lama dengan f_t untuk menghilangkan informasi yang tidak relevan, kemudian menambahkan dengan $\hat{C}_t$.
  $$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t$$
- Output gate: menghitung output.
  $$o_t = \sigma(x_t U^o + h_{t-1} W^o)$$
  $$h_t = \tanh(C_t) \cdot o_t$$

17

Bidirectional LSTM



- Terdiri atas double layer LSTM dengan arah yang berbeda
- Prediksi berdasarkan pada input sebelumnya dan input selanjutnya
- Perhitungan Bi-LSTM dengan L adalah layer **backward** LSTM, pada waktu t, dapat dituliskan pada persamaan berikut:

$$\overleftarrow{f}_t^L = \sigma\left(W_{\overleftarrow{f}h}^L h_{t+1}^L + W_{\overleftarrow{f}x}^L h_t^{L-1} + b_{\overleftarrow{f}}^L\right)$$

$$\overleftarrow{i}_t^L = \sigma\left(W_{i_h}^L h_{t+1}^L + W_{\overleftarrow{i}_x}^L h_t^{L-1} + b_i^L\right)$$

$$\overleftarrow{\tilde{c}}_t^L = \tanh\left(W_{\overleftarrow{\tilde{c}}h}^L h_{t+1}^L + W_{\overleftarrow{\tilde{c}}x}^L h_t^{L-1} + b_{\overleftarrow{\tilde{c}}}^L\right)$$

$$\overleftarrow{c}_t^L = \overleftarrow{f}_t^L \cdot \overleftarrow{\tilde{c}}_{t+1}^L + \overleftarrow{i}_t^L \cdot \overleftarrow{\tilde{c}}_t^L$$

$$\overleftarrow{o}_t^L = \sigma\left(W_{\overleftarrow{o}_h}^L h_{t+1}^L + W_{\overleftarrow{o}_x}^L h_t^{L-1} + b_{\overleftarrow{o}}^L\right)$$

$$\overleftarrow{h}_t^L = \overleftarrow{o}_t^L \cdot \tanh(\overleftarrow{c}_t^L)$$

$$y_t = W_{\vec{h}y}\vec{h}_t + W_{\vec{h}y}\overleftarrow{h}_t + b_y$$

19

Dataset

Dataset	Banyak Video	Format Video	Sumber
Hockey	1000	.avi	https://academictorrents.com/details/38d9ed996a5a75a039b84cf8a137be794e7cee89
Movies	200	.mpg, .mp4	https://academictorrents.com/details/70e0794e2292fc051a13f05ea6f5b6c16f3d3635
Crowd	246	.avi	https://www.openu.ac.il/home/hassner/data/violentflows/
Child	332	.mp4	Mengumpulkan video kekerasan pada anak dari berbagai platform online
Fight Surveillance Dataset	300	.mp4	Kumpulan video kejadian kekerasan dari rekaman CCTV dari LN

20

DATASET HOCKEY

21

CROWD DATASET

22

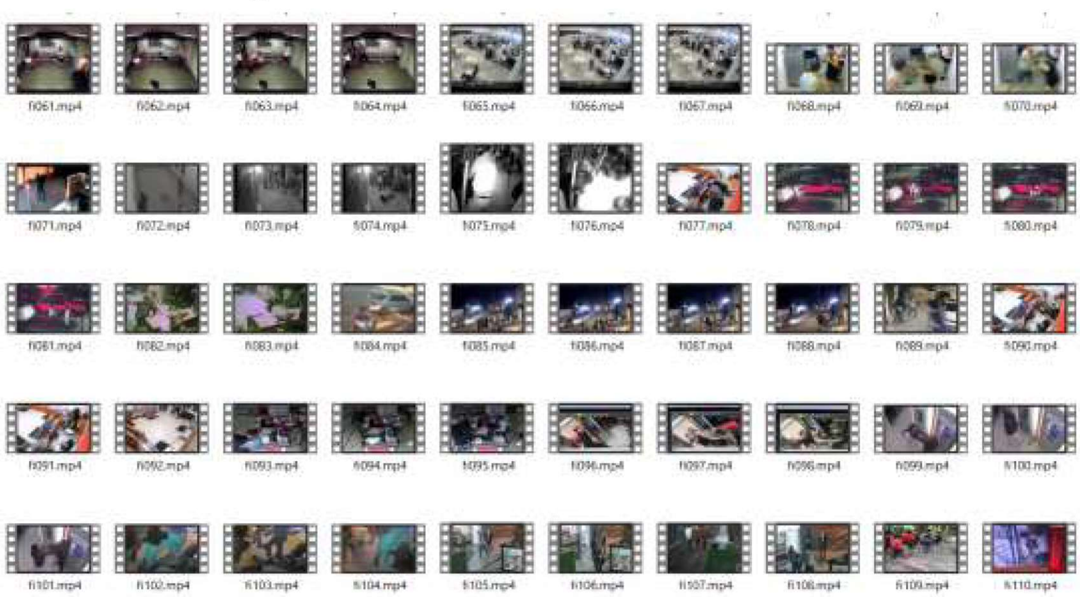
MOVIE DATASET

23

CHILD DATASET

24

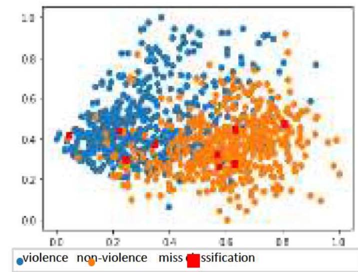
Fight Surveillance Dataset

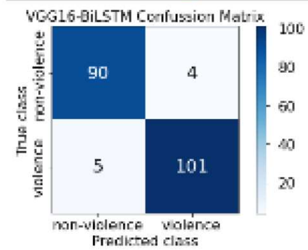


25

Hasil Ujicoba dengan Dataset Hockey

Classifier	F.Ext	Accuracy	Sensitifity	Specificity	G-Mean	Testing Time
SVM	VGG16	0.945	0.962	0.926	0.944	5.160
	Wavelet	0.950	0.962	0.936	0.949	29.512
	-	0.945	0.962	0.926	0.944	26.194
LSTM	VGG16	0.945	0.972	0.915	0.943	2.574
	Wavelet	0.940	0.915	0.968	0.941	7.049
	-	0.915	0.915	0.915	0.915	7.384
BiLSTM	VGG16	**0.955**	0.962	0.947	0.955	2.358
	Wavelet	0.930	0.906	0.957	0.931	13.605
	-	0.930	0.934	0.926	0.930	13.220
CNN-ResNet50[1]		0.875				
Inception Resnet V2[2]		0.933				
STACOG features + SVM[3]		0.904				





[1] M. Sharma and R. Baghel, *Video Surveillance for Violence Detection Using Deep Learning*, vol. 37. 2020.

[2] A. Jain and D. K. Vishwakarma, "Deep neuralNet for violence detection using motion features from dynamic images," *Proc. 3rd Int. Conf. Smart Syst. Inven. Technol. ICSSIT 2020*, no. Icssit, pp. 826–831, 2020, doi: 10.1109/ICSSIT48917.2020.9214153.

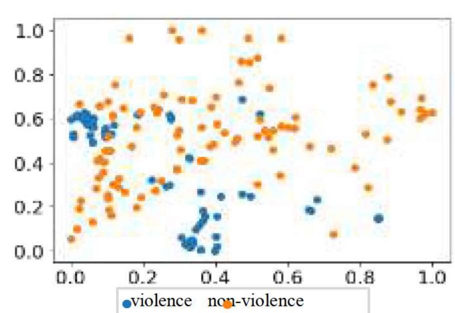
[3] K. Deepak, L. K. P. Vignesh, and S. Chandrakala, "Autocorrelation of gradients based violence detection in surveillance videos," *ICT Express*, vol. 6, no. 3, pp. 155–159, Sep. 2020, doi: 10.1016/J.ICTE.2020.04.014.

26

26

Hasil Ujicoba dengan Dataset Movie

Classifier	F.Ext	Accuracy	Sensitifity	Specificity	G-Mean	Testing Time
SVM	VGG16	1	1	1	1	0.052
	DWT	1	1	1	1	0.009
	-	1	1	1	1	0.461
LSTM	VGG16	1	1	1	1	0.698
	DWT	1	1	1	1	0.459
	-	1	1	1	1	0.707
BiLSTM	VGG16	1	1	1	1	1.323
	DWT	1	1	1	1	0.866
	-	1	1	1	1	1.209
Inception Resnet V2[1]		1				



[1]A. Jain and D. K. Vishwakarma, "Deep neuralNet for violence detection using motion features from dynamic images," *Proc. 3rd Int. Conf. Smart Syst. Inven. Technol. ICSSIT 2020*, no. Icssit, pp. 826–831, 2020, doi: 10.1109/ICSSIT48917.2020.9214153.

27

Kesimpulan

- Ujicoba dengan 5 dataset video yang berisi tindak kekerasan dan tidak, menunjukkan bahwa deep transfer learning khususnya VGG-16 meningkatkan performa sistem dalam mendeteksi tindak kekerasan.
- Deep transfer learning cocok digunakan untuk dataset dengan dimensi yang sangat besar, jumlah data yang sangat banyak, untuk sistem yang mengejar kecepatan dan juga akurasi.
- Pandangan yang perlu diperbaiki oleh sebagian besar kita adalah anggapan bahwa data latih dan data uji harus berasal dari sumber yang sama.

28

THANK YOU

29