Application of Attention Mechanism with Bidirectional Long Short-Term Memory (BiLSTM) and CNN for Human Conflict Detection using Computer Vision.

Aluno: Erick da Silva Farias

Orientador: Eduardo Palhares Júnior

Introduction

- Violence is a complex phenomenon that permeates the history of humanity.

Introduction

- Applications in monitoring and public safety policies.

Introduction

Dashdamirov, D. (2024). Comparative analysis: Violence recognition from videos using computer vision. 2024 IEEE 18th International Conference on Application of Information and Communication Technologies (AICT).
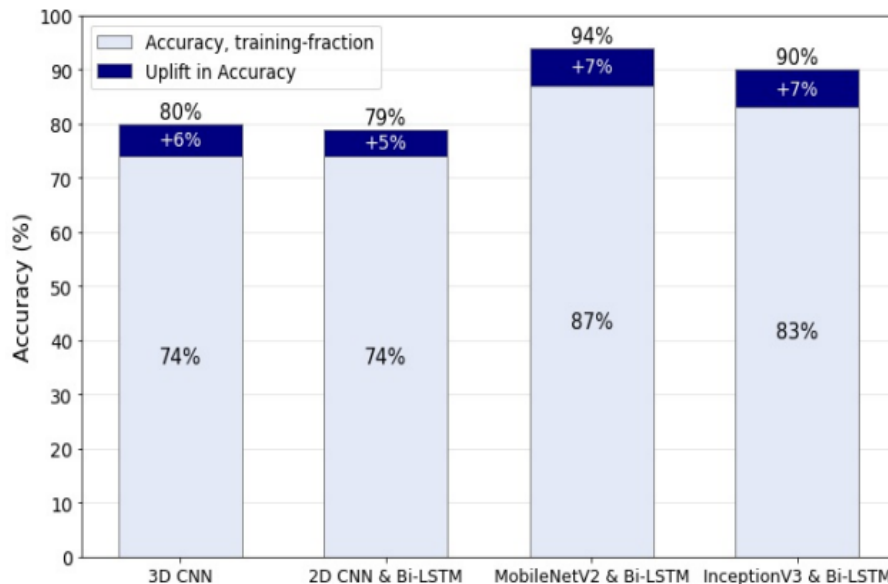
Dataset

- Real Life Violence Dataset

    - 1000 videos containing real street fight and 1000 video from other classes.

    - Videos from various sources.

Methodology

- Conversion the videos to sequence of frames

- Frames: 15

- Data augmentation: Scale, Multiply e Gaussian blur
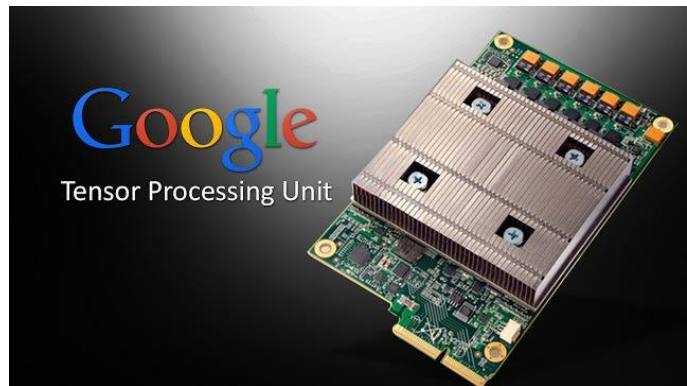
- HEIGHT and WIDTH: 100x100

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m1} & \cdots & a_{mn} \end{bmatrix}_{1n}^{} \begin{bmatrix} & & & \\ & & & \\ & & & \end{bmatrix}_{2n}^{1n}$$
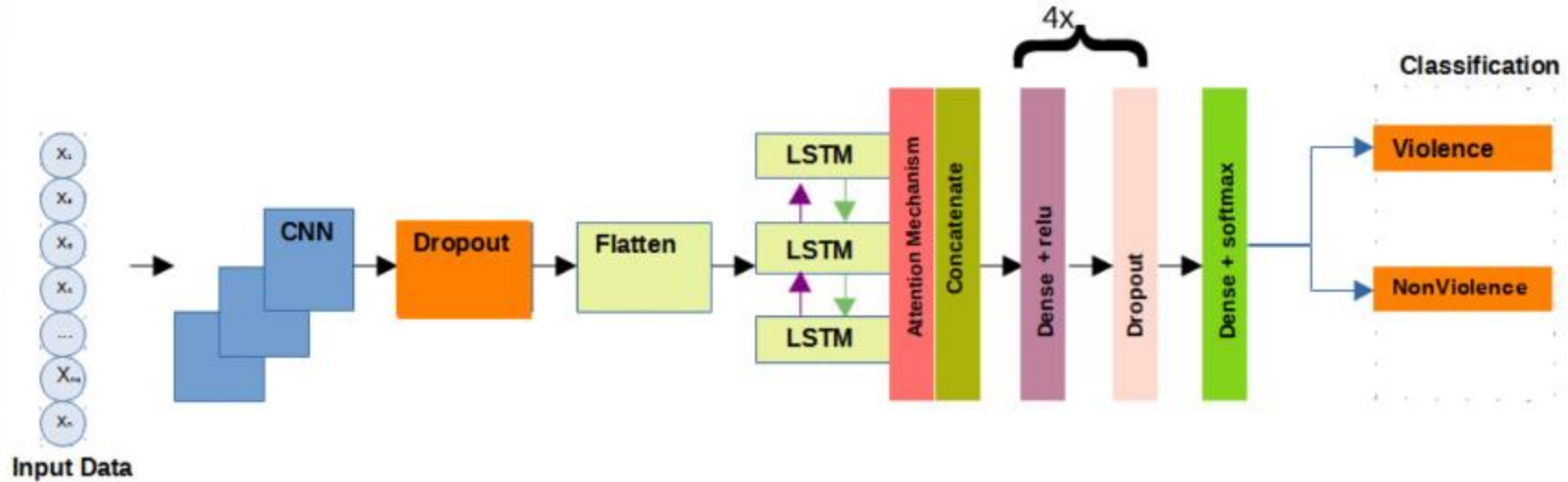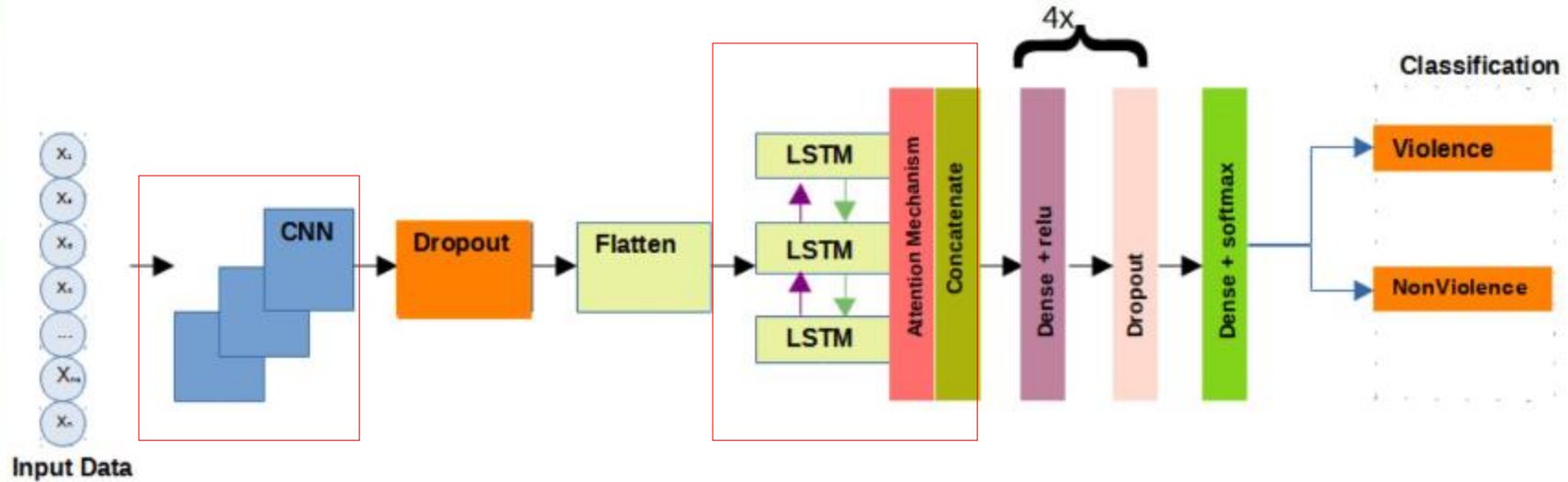
Methodology

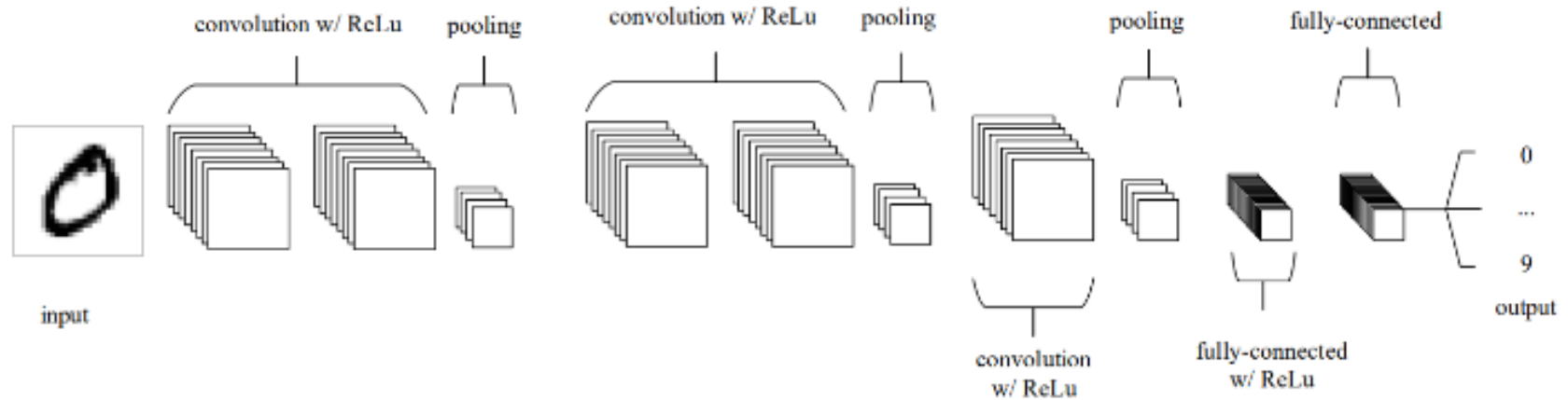- Transfer the dataset to the Cloud

Architecture implemented: CNN + BiLSTM + Mechanism Attention

Architecture implemented: CNN + BiLSTM + Mechanism Attention
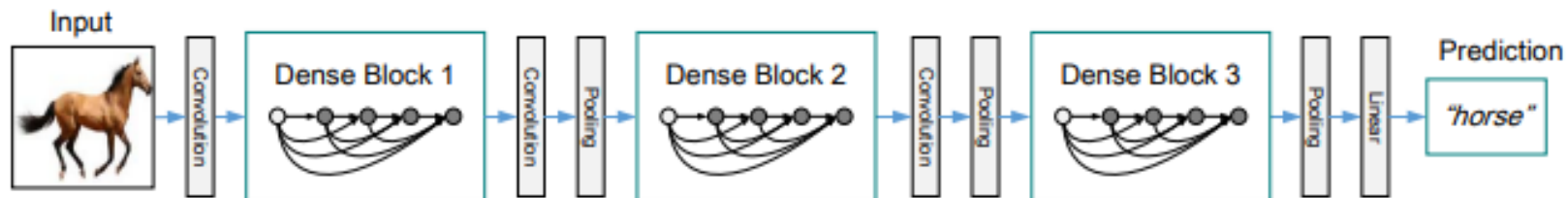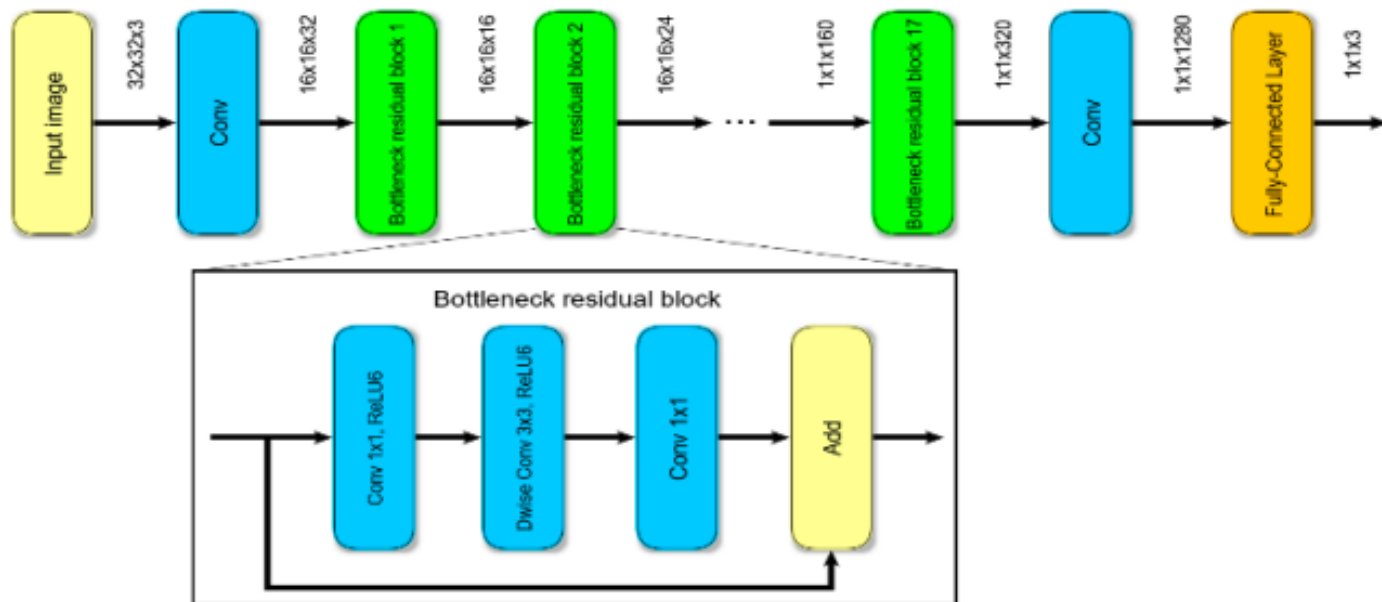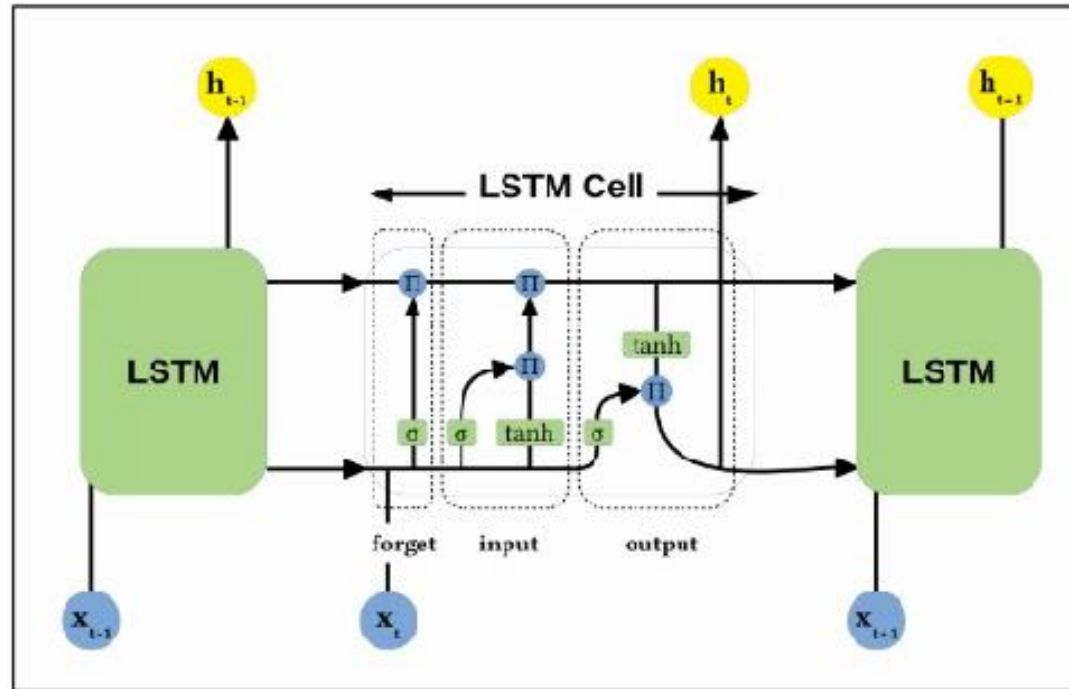
Convolutional Neural Network

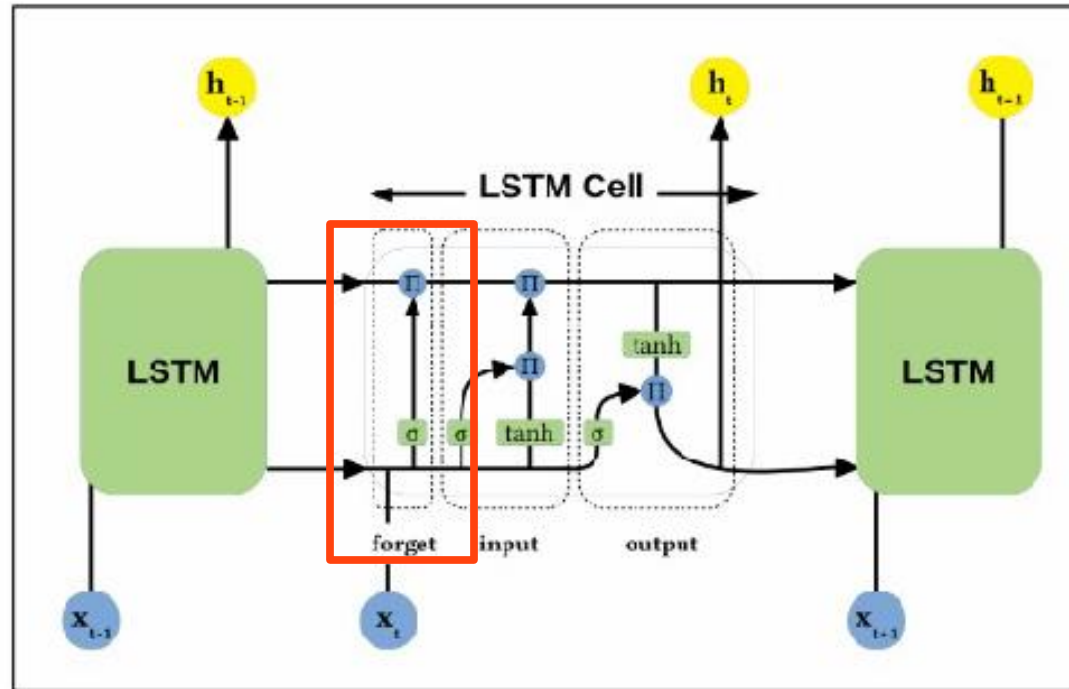# InceptionV3
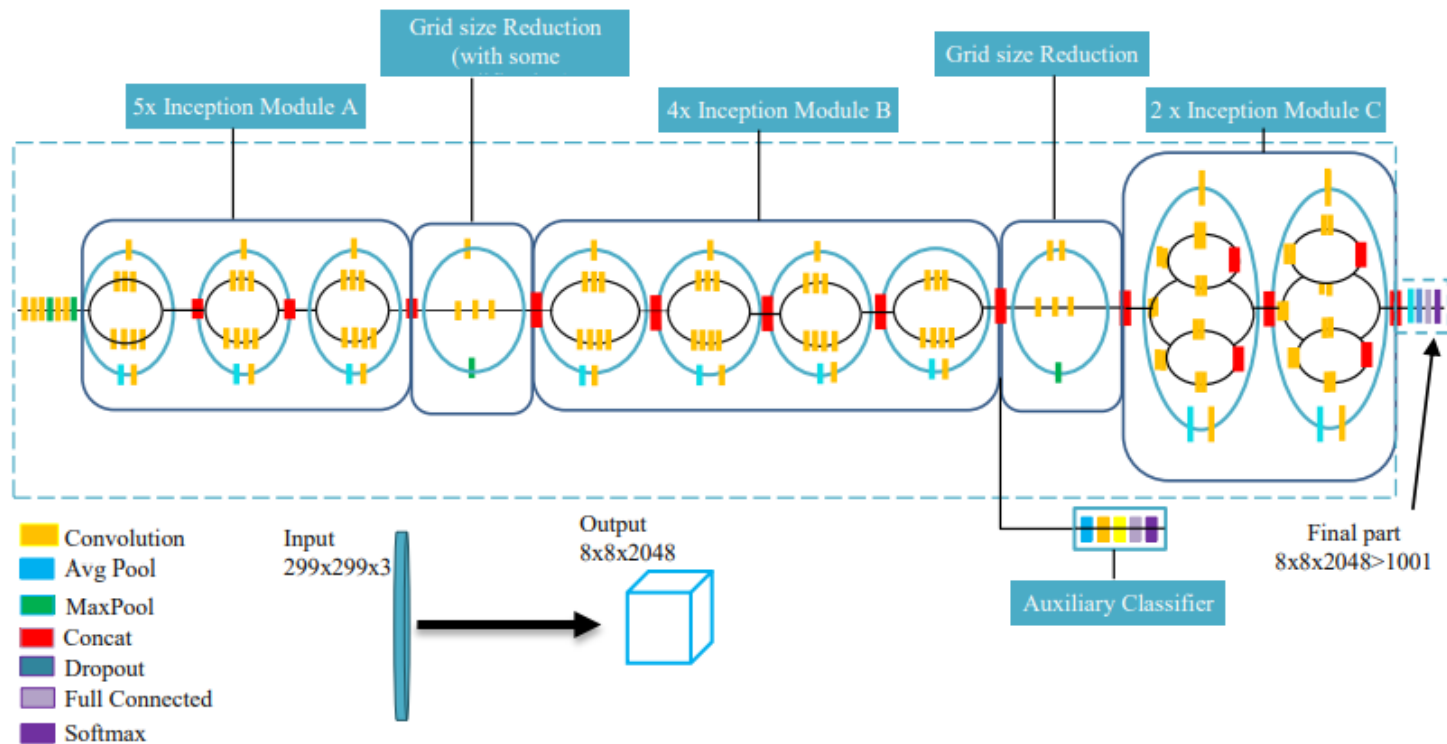
DenseNet121

MobileNetV2

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM)
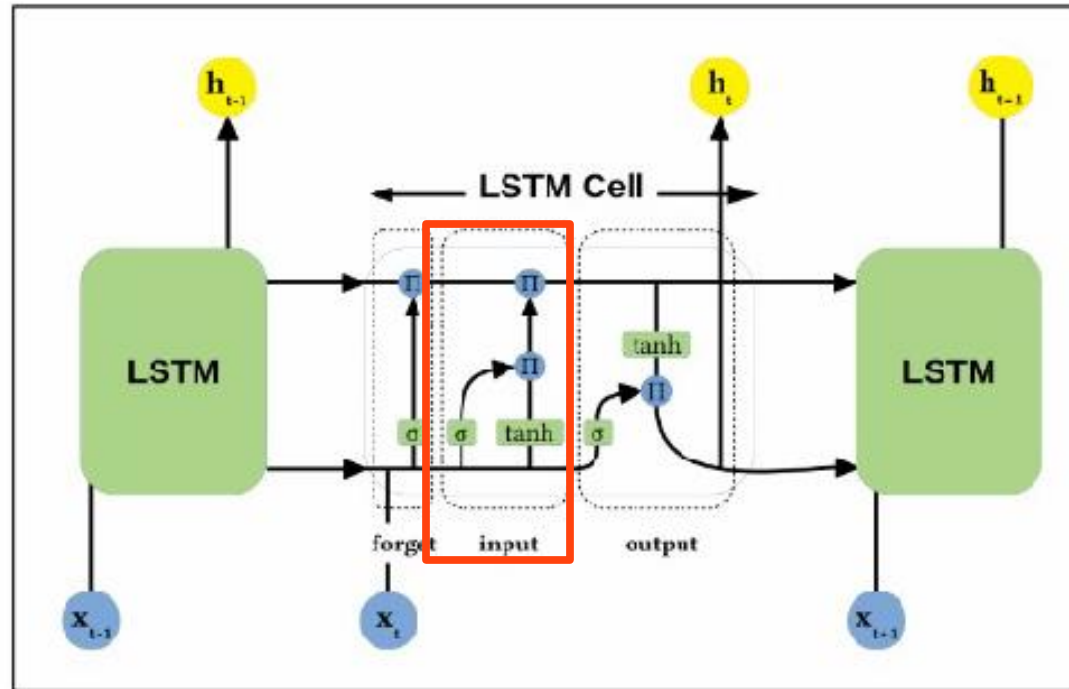
Long Short-Term Memory (LSTM)
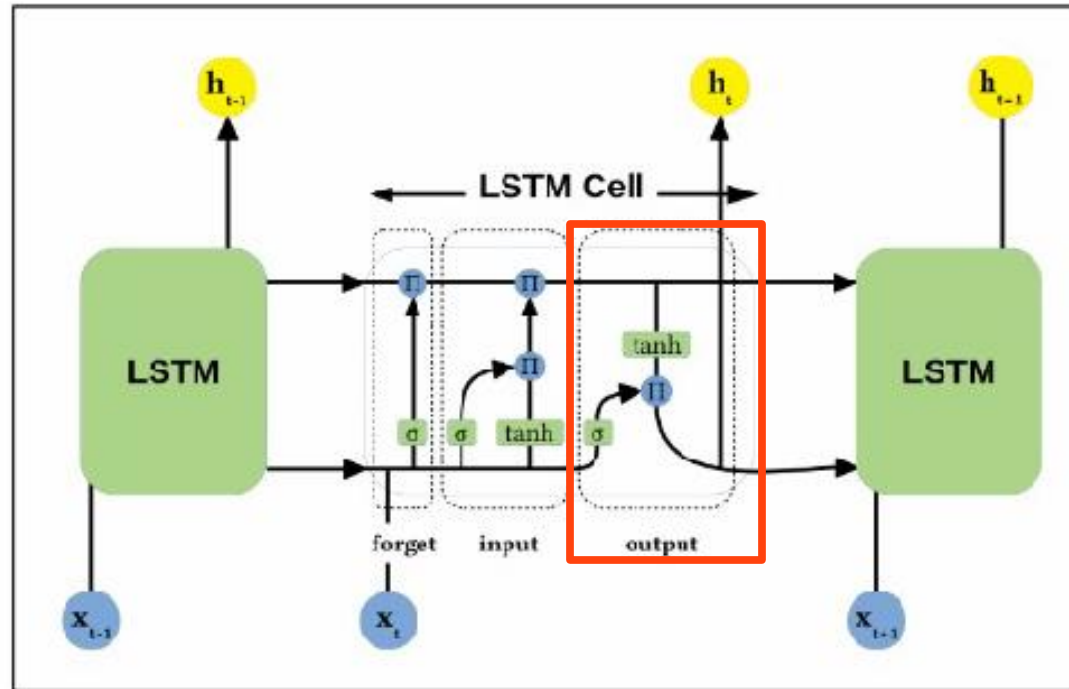
Bidirectional Long Short-Term Memory (BiLSTM)
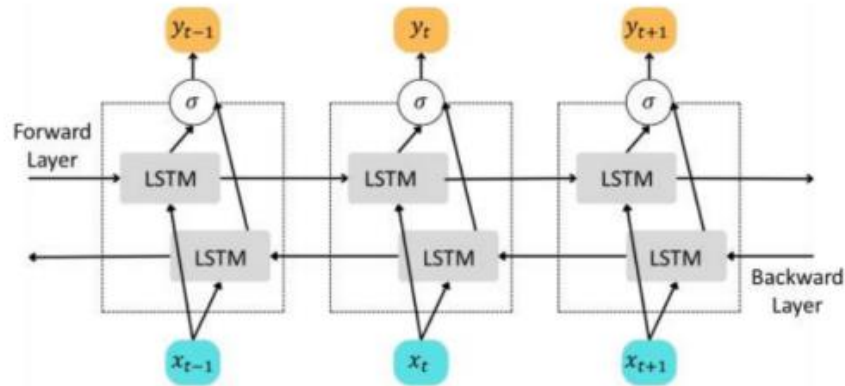
Mechanism Attention

$$e_i = \tanh(W^T x_i + b)$$

$$a_i = \frac{e_i}{\sum_j e_j}$$

$$output = \sum_i a_i x_i$$

Results

ID	Model	Attention	min_lr	batch_size	Accuracy (%)
1	MobileNetV2	No	0.0005	128	94.25
2	DenseNet121	No	0.0005	128	94.75
3	InceptionV3	No	0.0005	128	94.25
4	MobileNetV2	No	0.00005	64	89.00
5	DenseNet121	No	0.00005	64	93.75
6	InceptionV3	No	0.00005	64	91.00
7	MobileNetV2	Yes	0.0005	128	93.25
8	DenseNet121	Yes	0.0005	128	92.50
9	InceptionV3	Yes	0.0005	128	91.75
10	MobileNetV2	Yes	0.00005	64	96.50
11	DenseNet121	Yes	0.00005	64	95.50
12	InceptionV3	Yes	0.00005	64	94.25

Results

ID	Model	Attention	min_lr	batch_size	Accuracy (%)
1	MobileNetV2	No	0.0005	128	94.25
2	DenseNet121	No	0.0005	128	94.75
3	InceptionV3	No	0.0005	128	94.25
4	MobileNetV2	No	0.00005	64	89.00
5	DenseNet121	No	0.00005	64	93.75
6	InceptionV3	No	0.00005	64	91.00
7	MobileNetV2	Yes	0.0005	128	93.25
8	DenseNet121	Yes	0.0005	128	92.50
9	InceptionV3	Yes	0.0005	128	91.75
10	MobileNetV2	Yes	0.00005	64	96.50
11	DenseNet121	Yes	0.00005	64	95.50
12	InceptionV3	Yes	0.00005	64	94.25

Results

ID	Model	Attention	min_lr	batch_size	Accuracy (%)
1	MobileNetV2	No	0.0005	128	94.25
2	DenseNet12	No	0.0005	128	94.75
3	InceptionV3	No	0.0005	128	94.25
4	MobileNetV2	No	0.00005	64	89.00
5	DenseNet12	No	0.00005	64	93.75
6	InceptionV3	No	0.00005	64	91.00
7	MobileNetV2	Yes	0.0005	128	93.25
8	DenseNet12	Yes	0.0005	128	92.50
9	InceptionV3	Yes	0.0005	128	91.75
10	MobileNetV2	Yes	0.00005	64	96.50
11	DenseNet12	Yes	0.00005	64	95.50
12	InceptionV3	Yes	0.00005	64	94.25

Results

ID	Model	Attention	min_lr	batch_size	Accuracy (%)
1	MobileNetV2	No	0.0005	128	94.25
2	DenseNet121	No	0.0005	128	94.75
3	InceptionV3	No	0.0005	128	94.25
4	MobileNetV2	No	0.00005	64	89.00
5	DenseNet121	No	0.00005	64	93.75
6	InceptionV3	No	0.00005	64	91.00
7	MobileNetV2	Yes	0.0005	128	93.25
8	DenseNet121	Yes	0.0005	128	92.50
9	InceptionV3	Yes	0.0005	128	91.75
10	MobileNetV2	Yes	0.00005	64	96.50
11	DenseNet121	Yes	0.00005	64	95.50
12	InceptionV3	Yes	0.00005	64	94.25

Results

ID	Model	Attention	min_lr	batch_size	Accuracy (%)
1	MobileNetV2	No	0.0005	128	94.25
2	DenseNet121	No	0.0005	128	94.75
3	InceptionV3	No	0.0005	128	94.25
4	MobileNetV2	No	0.00005	64	89.00
5	DenseNet121	No	0.00005	64	93.75
6	InceptionV3	No	0.00005	64	91.00
7	MobileNetV2	Yes	0.0005	128	93.25
8	DenseNet121	Yes	0.0005	128	92.50
9	InceptionV3	Yes	0.0005	128	91.75
10	MobileNetV2	Yes	0.00005	64	96.50
11	DenseNet121	Yes	0.00005	64	95.50
12	InceptionV3	Yes	0.00005	64	94.25

Results

ID	Model	Attention	min_lr	batch_size	Accuracy (%)
1	MobileNetV2	No	0.0005	128	94.25
2	DenseNet121	No	0.0005	128	94.75
3	InceptionV3	No	0.0005	128	94.25
4	MobileNetV2	No	0.00005	64	89.00
5	DenseNet121	No	0.00005	64	93.75
6	InceptionV3	No	0.00005	64	91.00
7	MobileNetV2	Yes	0.0005	128	93.25
8	DenseNet121	Yes	0.0005	128	92.50
9	InceptionV3	Yes	0.0005	128	91.75
10	MobileNetV2	Yes	0.00005	64	96.50
11	DenseNet121	Yes	0.00005	64	95.50
12	InceptionV3	Yes	0.00005	64	94.25

Results

| ID	Model	Attention	min_lr	batch_size	Accuracy (%)
1	MobileNetV2	No	0.0005	128	94.25
2	DenseNet121	No	0.0005	128	94.75
3	InceptionV3	No	0.0005	128	94.25
4	MobileNetV2	No	0.00005	64	89.00
5	DenseNet121	No	0.00005	64	93.75
6	InceptionV3	No	0.00005	64	91.00
7	MobileNetV2	Yes	0.0005	128	93.25
8	DenseNet121	Yes	0.0005	128	92.50
9	InceptionV3	Yes	0.0005	128	91.75
10	MobileNetV2	Yes	0.00005	64	96.50
11	DenseNet121	Yes	0.00005	64	95.50
12	InceptionV3	Yes	0.00005	64	94.25

Results

Results
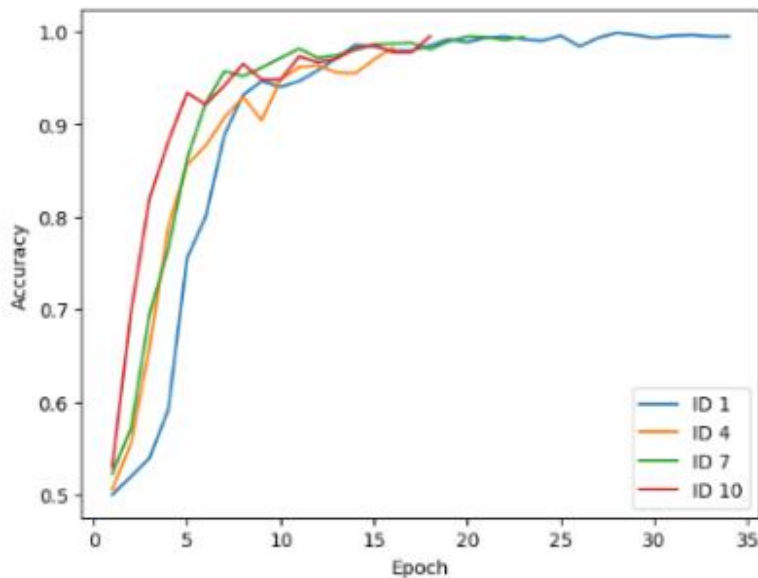
Accuracy in relation to the epochs of all training of the model MobileNetV2.
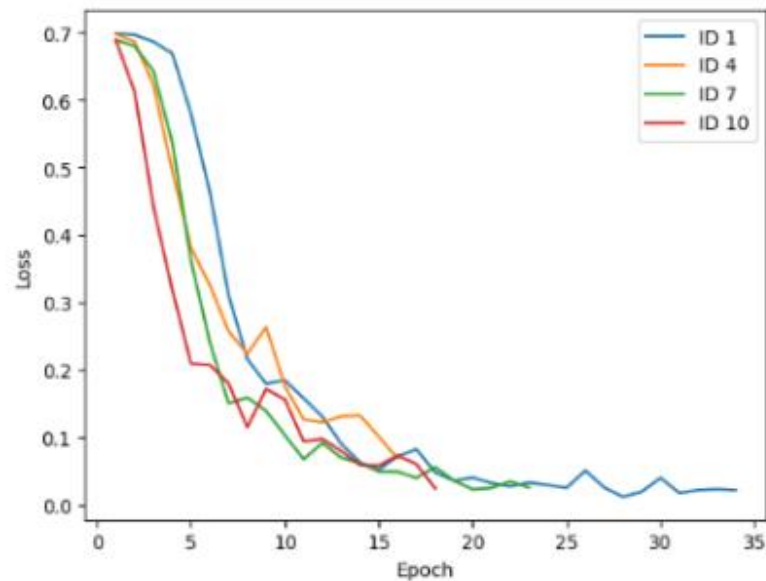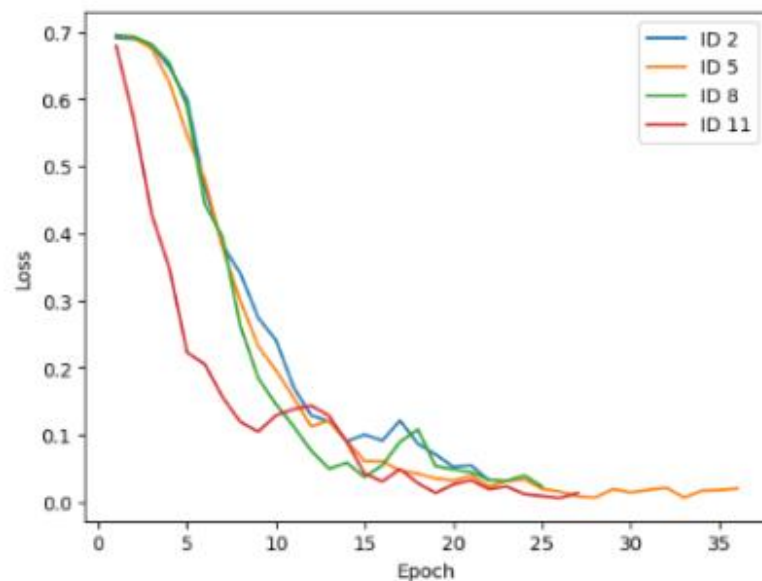

Error in relation to the epochs of all training of the model MobileNetV2
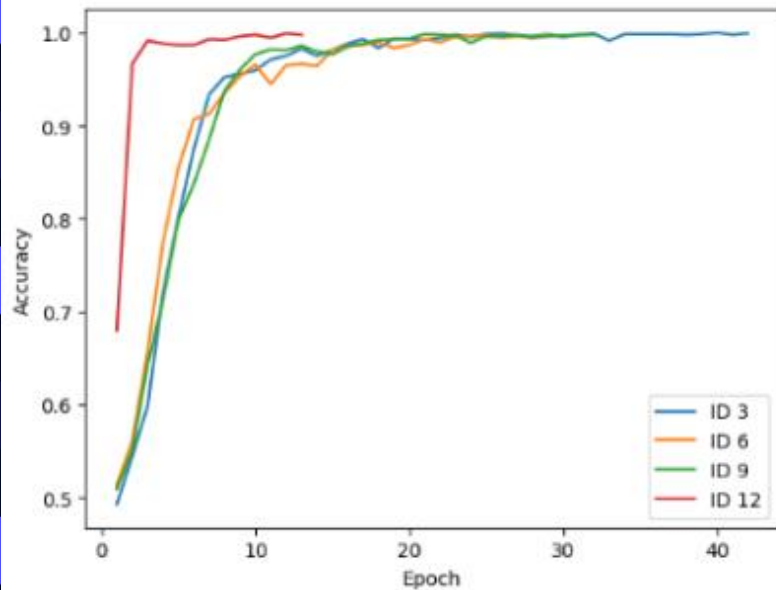
Results
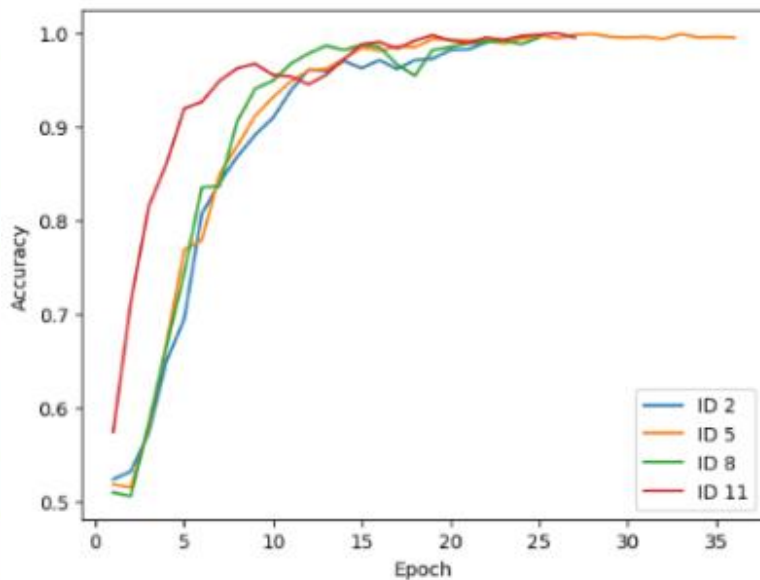
Accuracy in relation to the epochs of all training of the model DenseNet121
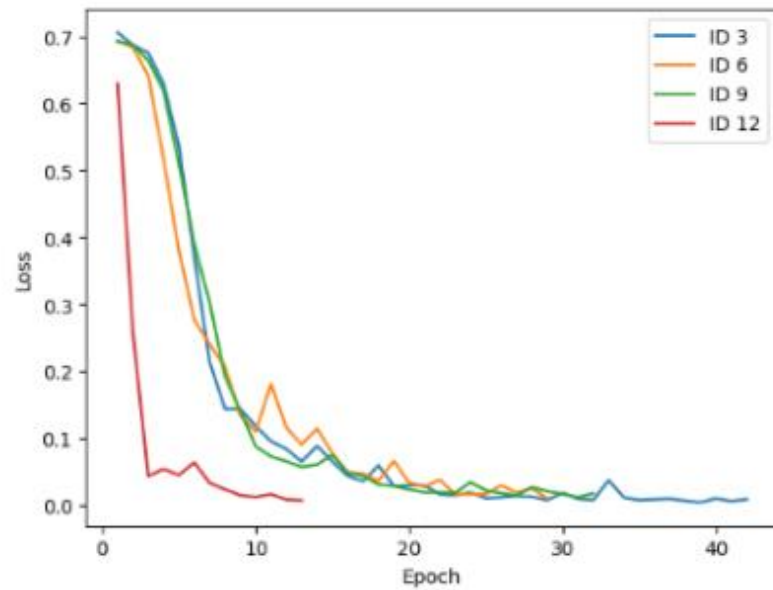

Error in relation to the epochs of all training of the model DenseNet121

Results

Accuracy in relation to the epochs of all training of the model InceptionV3

Error in relation to the epochs of all training of the model InceptionV3

Conclusion

Main Results:

- All models were effective.
- MobileNetV2 stood out, achieving the highest accuracy (96.50%), especially with reduced batch size and the application of the attention mechanism.

Importance of Parameters:

- The appropriate selection of parameters, such as learning rate and batch size, is crucial to optimize the models' performance.

- The attention mechanism showed a slight reduction in accuracy in some scenarios but proved effective in specific parameter combinations (for example, learning rate of 0.00005 and batch size of 64).

Suggestions for Future Work:

- Explore the use of multimodal data (combining video, audio, images, and sensor information).

- Investigate how integrating multiple sources can lead to a more robust scene analysis.

AGRADECIMENTO