Fall 11-15-2021

How Facial Features and Head Gesture Convey Attention in Stationary Environments

Janelle Domantay
University of Nevada, Las Vegas, domantay@unlv.nevada.edu

Brendan Morris Ph.D.
University of Nevada, Las Vegas, brendan.morris@unlv.edu

HOW FACIAL FEATURES CONVEY ATTENTION IN STATIONARY ENVIRONMENTS

Janelle Domantay

RESEARCH QUESTION

Using raw video data, what method can best predict whether the subject is alert or drowsy?

HOW WE ANALYZE

Support Vector Machines (SVM)

- Takes in extracted features in the form of numeric values
- Outputs a prediction

OPENFACE

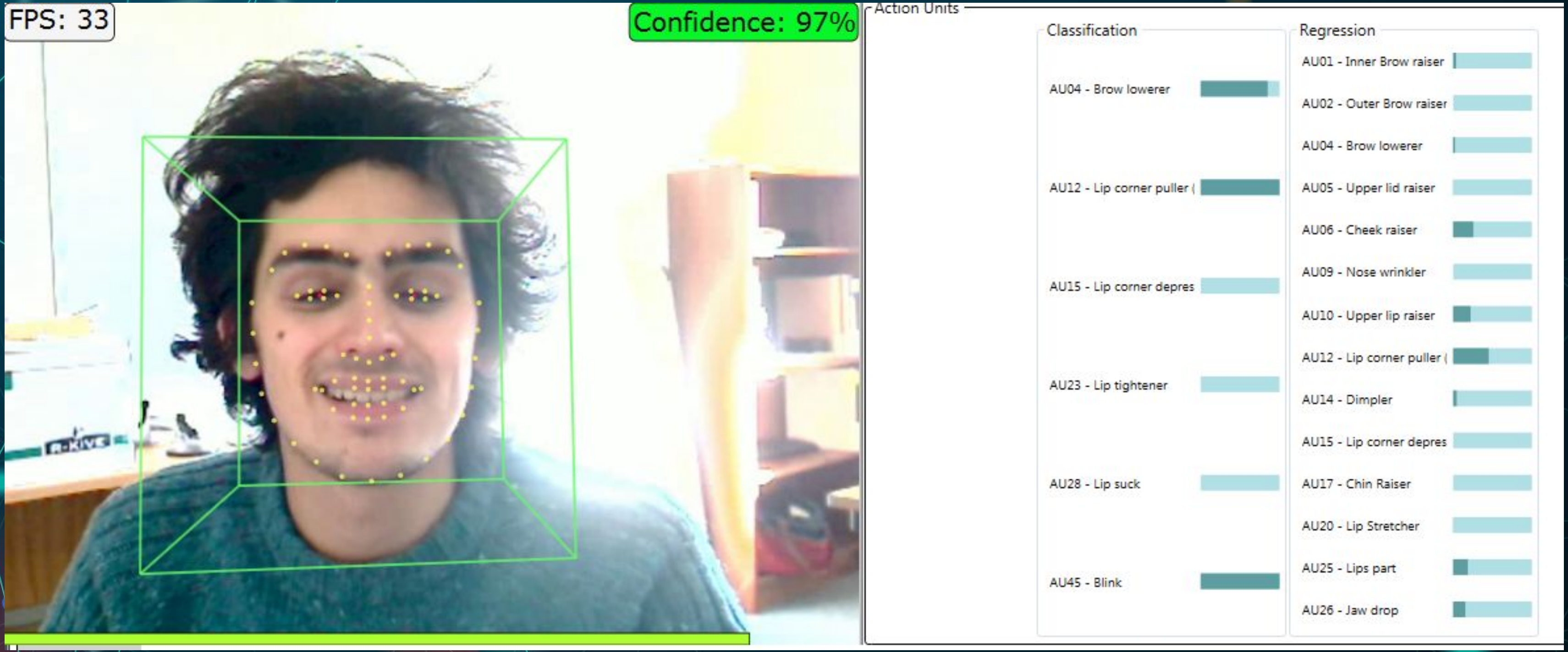


Figure 4: Sample Images of OpenFace analysis toolkit from Baltrusaitis Github [2]

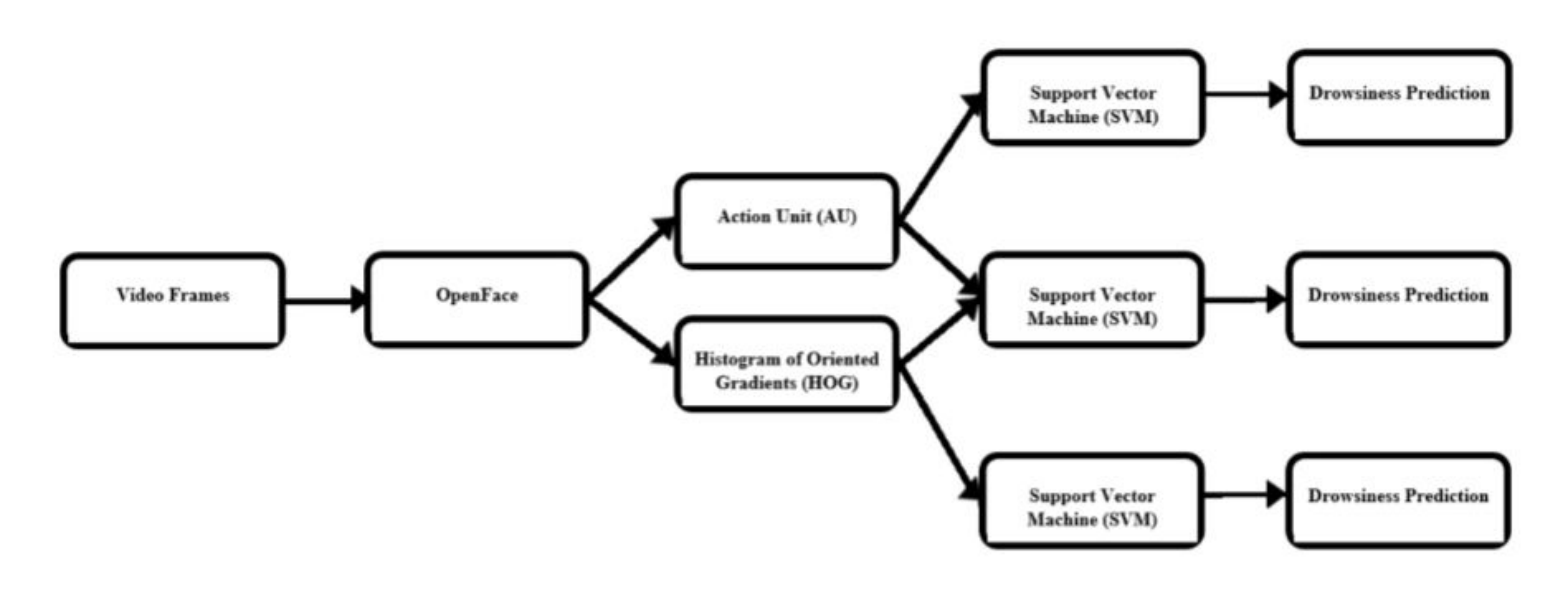


Figure 1: Support Vector Machine Diagram

HOW WE ANALYZE

Support Vector Machines (SVM)

- Takes in extracted features in the form of numeric values
- Outputs a prediction

Convolutional Neural Network

- Takes in a still image
- Outputs a prediction

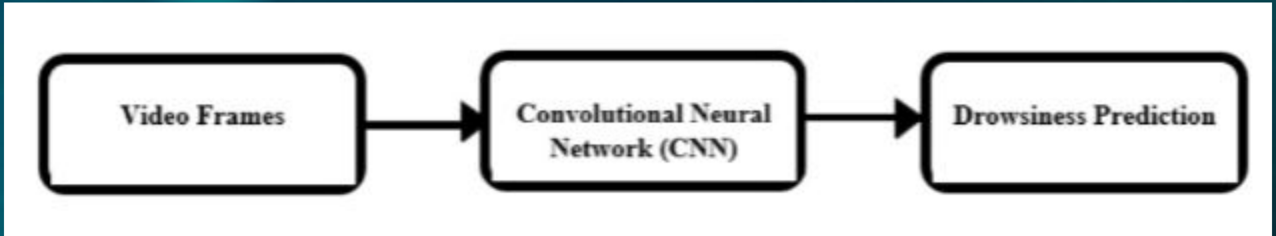


Figure 2: Convolutional Neural Network Diagram

HOW WE ANALYZE

Support Vector Machines (SVM)

- Takes in extracted features in the form of numeric values
- Outputs a prediction

Convolutional Neural Network

- Takes in a still image
- Outputs a prediction

Convolutional Recurrent Neural Network

- Takes in a video sequence:
    - Feeds still image to CNN
    - Feeds CNN to Recurrent Neural Network
- Outputs a prediction

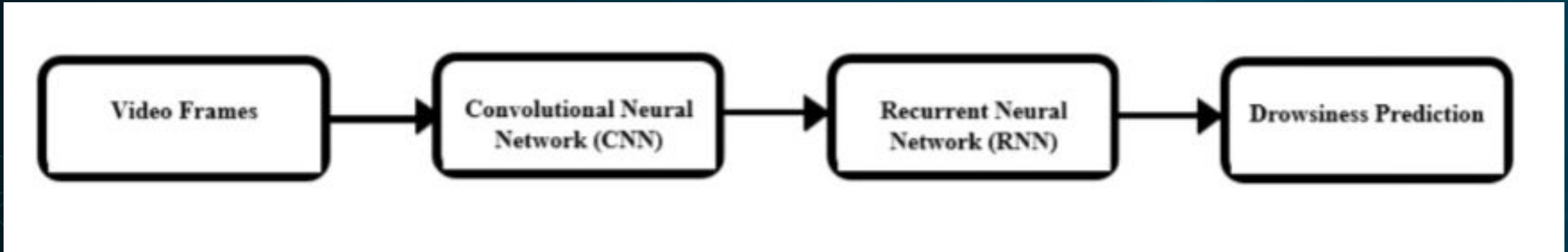


Figure 3: Recurrent Neural Network Diagram

HOW WE PREDICT

1. Data is divided into TRAINING DATA and TEST DATA
2. A model makes predictions on TRAINING DATA and is penalized for incorrect predictions
3. A model makes changes to itself
4. A model is evaluated on Test Data

Methodology & Results

Experimental Procedure

WHAT IS ATTENTION?

Attention = Not Drowsy

University of Texas at Arlington Real-Life Drowsiness Dataset (UTA-RLDD)

1-Extremely Alert
2-Very Alert
3-Alert
4-Rather Alert
5-Neither Alert or Sleepy
6-Some signs of sleepiness
7-Sleepy, but no difficulty remaining awake
8-Sleepy, some effort to keep alert
9-Extremely Sleepy, fighting sleep

Figure 5: UTA-RLDD drowsiness scale [1].

DATA COLLECTION

Figure 6: Sample frames from Ghoodosian's [4] UTA Real-Life Drowsiness Dataset, alert (first row), low vigilance (second row), drowsy (third row)

- UTA-RLDD [1]
  - 30 hours of 60 individuals
- 3000 image subset
  - 1000 images per class
- 28 frame video segments
- Training Set: before 7 minutes
  - ~800 frames
- Test Set: after 7 minutes
  - ~200 frames

SUPPORT VECTOR MACHINE

Attributes	Kernel	Precision	Recall	Test Acc.	Time(ms)
AU	RFC*	0.4685	0.4732	0.4732	4.549E-3
	Linear	0.4719	0.4685	0.4685	0.02006
	Polynomial	0.6762	0.6282	0.6282	0.01753
	Sigmoid	0.3747	0.3753	0.3753	0.03867
	Gaussian	0.7144	0.7145	0.7145	0.04795E
HOG	RFC	0.6291	0.6270	0.6270	0.1073E-4
	Linear	0.9067	0.9068	0.9068	3.812
	Polynomial	0.9551	0.9545	0.9545	4.975
	Sigmoid	0.5076	0.4848	0.4848	6.372
	Gaussian	0.9307	0.9301	0.9301	14.98
HOG & AU	RFC	0.6034	0.6014	0.6014	0.01099
	Linear	0.9103	0.9103	0.9103	3.897
	Polynomial	0.9563	0.9557	0.9557	5.102
	Sigmoid	0.4660	0.4545	0.4545	6.409
	Gaussian	0.9271	0.9266	0.9266	15.388

Table 1: Support Vector Machine Results

- Best Performance:
  - HOG & AU
  - Accuracy: 95.57 %
  - Polynomial
- Overall Performance: Varied

SUPPORT VECTOR MACHINE

Attributes	Kernel	Precision	Recall	Test Acc.	Time(ms)
AU	RFC	0.4095	0.4103	0.4103	4.822E-3
	Linear	0.3520	0.3504	0.3520	0.01348
	Polynomial	0.4563	0.4219	0.4219	0.01106
	Sigmoid	0.3716	0.3660	0.3660	0.02695
	Gaussian	0.4601	0.4580	0.4580	0.03235
HOG	RFC	0.6104	0.6010	0.6011	0.1057
	Linear	0.5291	0.5245	0.5245	1.651
	Polynomial	0.9431	0.9429	0.9429	1.312
	Sigmoid	0.9282	0.9277	0.9277	1.688
	Gaussian	0.4815	0.4464	0.4464	2.326
HOG & AU	RFC	0.6770	0.6760	0.6760	0.1040
	Linear	0.9488	0.9487	0.9487	1.663
	Polynomial	0.9604	0.9604	0.9604	1.264
	Sigmoid	0.5045	0.4709	0.4709	1.671
	Gaussian	0.9467	0.9464	0.9464	2.337

Table 2: Support Vector Machine Results using Video

- ❖ Best Performance:
  - ➢ HOG & AU
  - ➢ Accuracy: 96.04 %
  - ➢ Polynomial
- ❖ Overall Performance: Varied

CONVOLUTIONAL NEURAL NETWORK

CNN	Epochs	Acc.	Loss	Val.	V. Loss	Time(s)
MobileNetV2	100	0.9993	0.0041	0.6695	1.0328	1.955
ResNet50	100	0.9980	0.0064	0.9880	0.0355	2.131
DenseNet121	100	1.00	0.0057	0.9896	0.0453	2.866
InceptionV3	100	0.9973	0.0055	0.9964	0.0142	2.560

Table 3: CNN results from scratch

CNN	Epochs	Acc.	Loss	Val.	V. Loss	Time(s)
MobileNetV2	20	0.9983	0.0110	0.9916	0.0308	2.007
ResNet50	20	0.9985	0.0088	0.9928	0.0192	2.025
DenseNet121	20	0.9992	0.0134	0.9784	0.0634	3.354
InceptionV3	20	0.9960	0.0279	0.9892	0.0441	2.581

Table 4: CNN results via transfer learning

- Best Performance:
  - InceptionV3
  - Acc: 99.64 %
- Best Performance:
  - ResNet50
  - Acc: 99.28 %

CONVOLUTIONAL RECURRENT NEURAL NETWORK

CRNN	Epochs	Acc.	Loss	Val.	Val. Loss	Time(s)*
CRNN	100	1.000	1.745E-4	0.9697	0.2158	0.05449
ResNetCRNN	100	1.000	7.227E-4	0.9848	0.0992	14.36
3DCNN	15	1.000	0.0201	0.9545	0.1609	0.2814

*per 28 frame sequence

Table 5: CRNN results

- Best Performance:
  - ResNetCRNN
  - Acc: 98.48 %

QUANTITATIVE EVALUATION

Method	SVM	SVM (V)	CNN (S)	CNN (TL)	RNN
Best Performance	HOG & AU Polynomial Kernel 95.57 % Accuracy	HOG & AU Polynomial Kernel 96.04% Accuracy	InceptionV3 99.64 % Accuracy	ResNet50 99.28 % Accuracy	ResNetCRNN 98.48% Accuracy
Worst Performance	AU Sigmoid 37.53 % Accuracy	AU Random Forest Classifier 41.03 % Accuracy	MobileNetV2 66.95 % Accuracy	DenseNet121 97.84%	3DCNN 95.45 % Accuracy

Table 6: Quantitative evaluation

QUALITATIVE EVALUATION

- SVMs CAN approach high accuracy
- CNNs performed better than CRNNs
- SVMs require less processing power and less time
- SVMs explicitly analyze facial features

CHALLENGES AND FUTURE DIRECTIONS

- Models require some familiarity with a subject
- Studying additional attributes that convey attention

CONCLUSIONS AND APPLICATIONS

- SVMs require **low processing power**, can approach the **accuracy** of deep learning methods and can be employed in **real time**
- Open Source Attention Detection
  - Driver Drowsiness
  - Work Environments
  - Virtual Classrooms

REFERENCES

[1] T. Baltrušaitis, OpenFace 2.2.0: a facial behavior analysis toolkit [Source code]. 2020. OpenFace/README.md at master · TadasBaltrusaitis/OpenFace · GitHub.

[2] R. Ghoddoosian, M. Galib and V. Athitsos, "A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA

[3] T. Berrú, M. Yeferson & M. Palma. "Aplicación móvil para la detección de somnolencia de un conductor aplicando visión artificial," 2016.

[4] A. Singh. "Feature Engineering for Images: A valuable Introduction to the HOG Feature Descriptor," 2019. analytics vidhya.

[5] T. Gerard Lynn, P. Takako Endo, P. Rosati, and I. Silva. "A Comparison of Machine Learning Approaches for Detecting Misogynistic Speech in Urban Dictionary," 2019 IEEE Cyber Science 2019, Oxford, England

[6] S. Saha, A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. 2018. towards data science.

[7] C. Church Chatterjee. "Implementation of RNN, LSTM, and GRU," 2019. towards data science.

THANK YOU!

Do you have any questions?

domantay@unlv.nevada.edu

Additional Slides

ATTENTION METRICS

- Facial Expressions
- Percent Closure
- Histogram of Oriented Gradients (Edge Detection)

FACIAL EXPRESSION

Action Unit	Facial Feature
AU01	Inner brow raiser
AU02	Outer brow raiser
AU04	Brow lowerer
AU05	Upper lid raiser
AU06	Cheek raiser
AU07	Lid tightener
AU09	Nose wrinkler
AU10	Upper lip raiser
AU12	Lip corner puller
AU14	Dimpler
AU15	Lip corner depressor
AU17	Chin raiser
AU20	Lip stretcher
AU23	Lip tightener
AU25	Lips part
AU26	Jaw drop
AU28*	Lip Suck
AU45	Blink

*Can only be detected by presence (0 or 1)

Table 6: Action Units detected by OpenFace

- Facial Action Coding System (FACS)
  - Developed by an anatomist
  - Utilized by psychologists and animators
- Action Units

PERCLOS

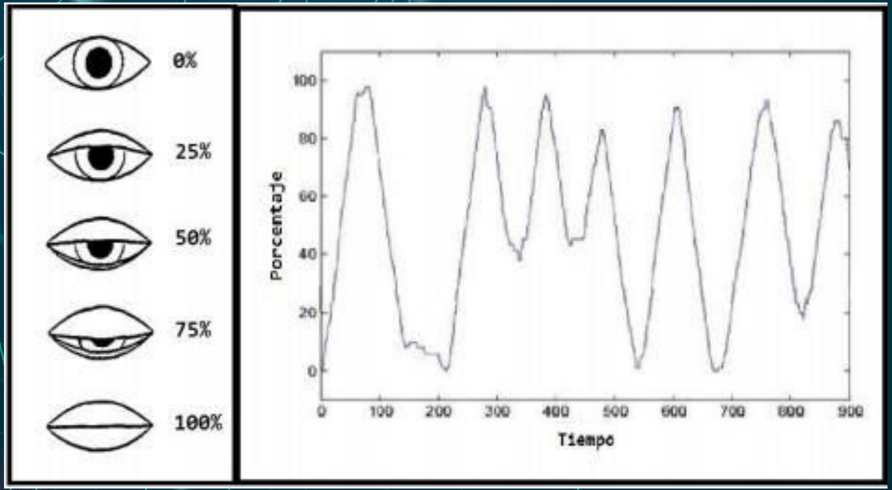


Figure 7: Visualization of PERCLOS tracking [3].

- PERCLOS (Percent Closure)
- “First Ever” real-time drowsiness detection sensor

HISTOGRAM OF ORIENTED GRADIENTS

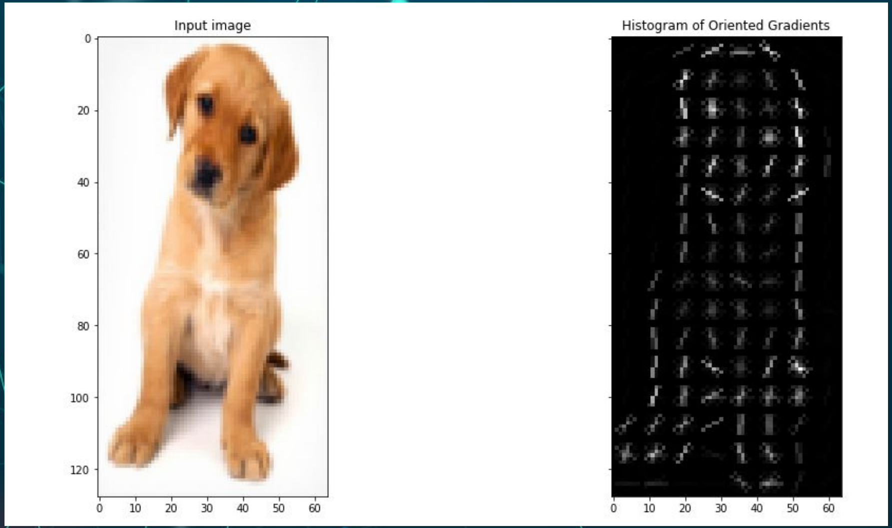


Figure 8: Visualization of HOG Feature Descriptor [4].

- ❖ Edge and orientation detection
- ❖ HOG

SUPPORT VECTOR MACHINES

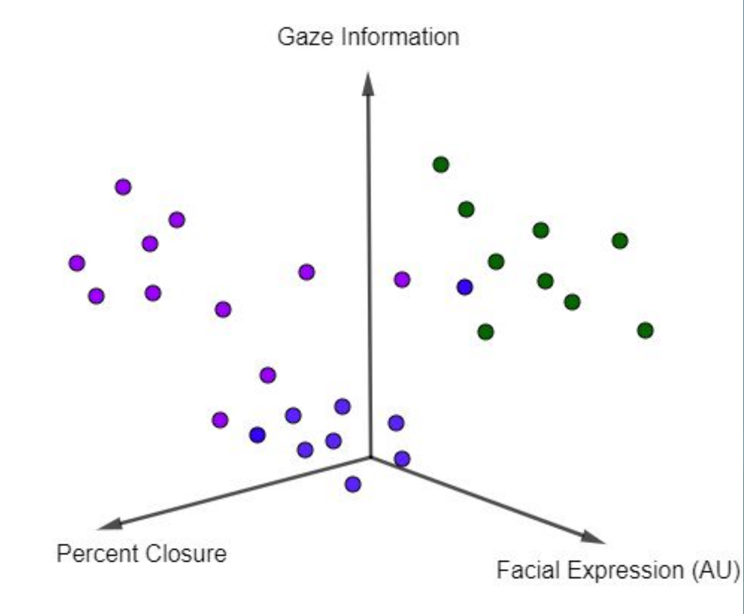


Figure 10: Example visualization of Training Data.

- Calculate the boundary between data values
- Requires explicit feature extraction

Green = Alert
Magenta = Fatigued
Purple = Distracted

SUPPORT VECTOR MACHINES

Figure 10: Example visualization of Training Data.

- Calculate the boundary between data values
- Requires explicit feature extraction
- Different **Kernel functions** can affect how boundaries are calculated

Green = Alert
Magenta = Fatigued
Purple = Distracted

SUPPORT VECTOR MACHINES

True Positive = it belongs to the class and was classified into it	False Positive = it does not belong to the class and was classified into it
False Negative = it belongs to the class and was not classified into it	True Negative = it does not belong to the class and was not classified into it

Figure 11: Possible categories a classification can fall into.

- Evaluation Metrics
  - Accuracy =
    - (TP + TN) / (TP + FP + FN + TN)
  - Precision
    - TP / (TP + FP)
  - Recall
    - TP / (TP + FN)
  - Time (ms)

DEEP LEARNING: NEURAL NETWORKS

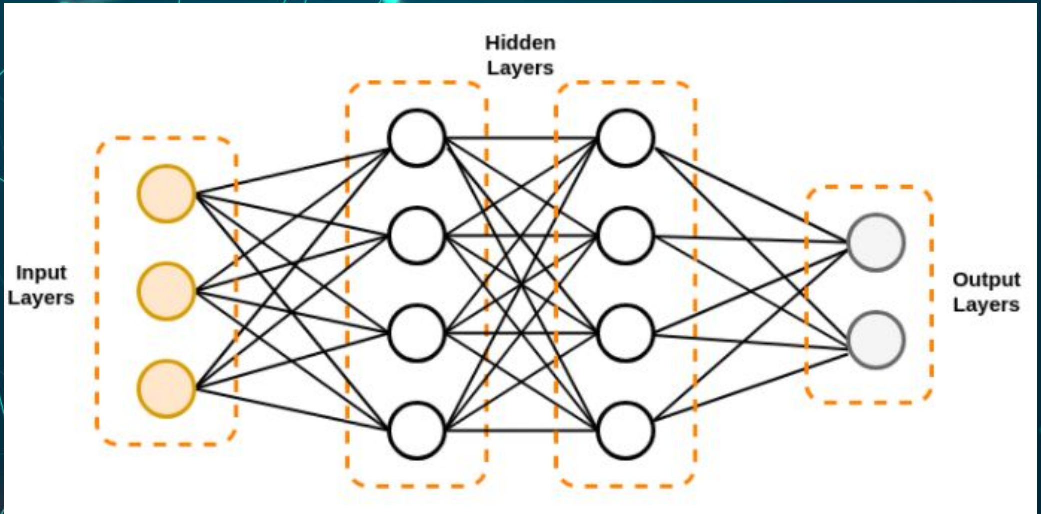
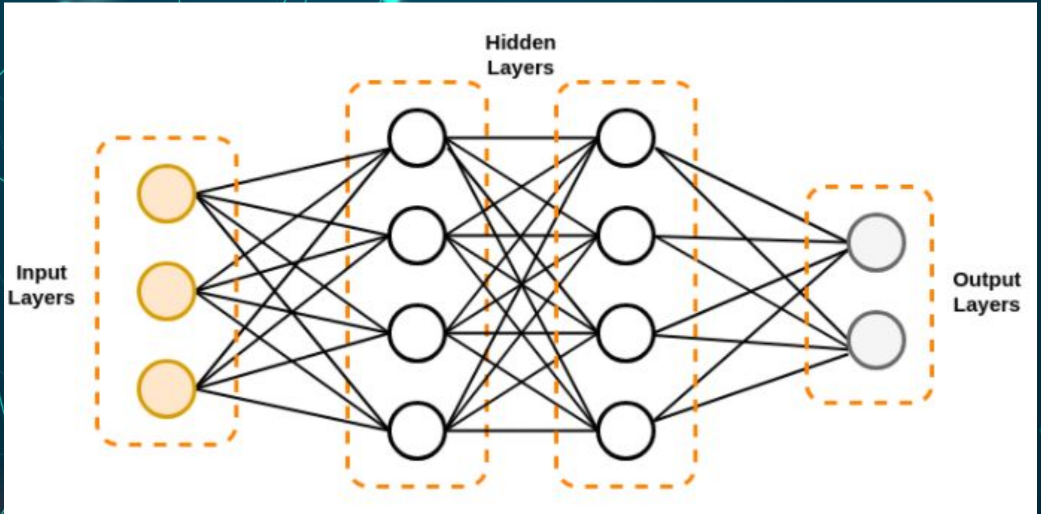


Figure 12: Image of generic neural network [5]

- Transfer Learning
- Epochs
- Loss
- Time (s)

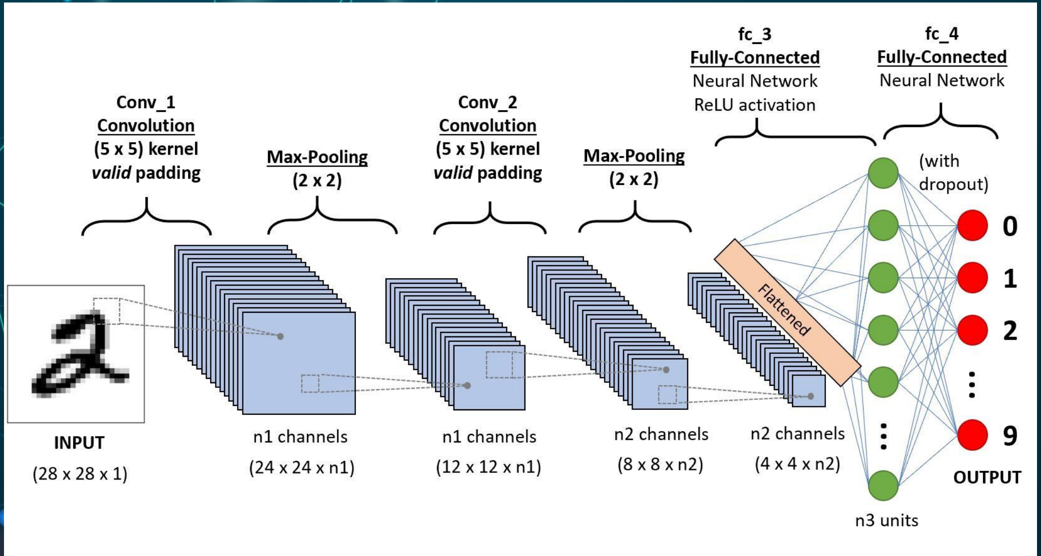
DEEP LEARNING: NEURAL NETWORKS



- ❖ Transfer Learning
- ❖ Higher Epochs
- ❖ Lower Loss

Figure 12: Image of generic neural network [3]

DEEP LEARNING: NEURAL NETWORKS



- Convolutional Neural Networks (CNN)
- 3D Convolutional Neural Network (3DCNN)

Figure 13: Image of CNN sequence for MNIST classification [6]

DEEP LEARNING: NEURAL NETWORKS

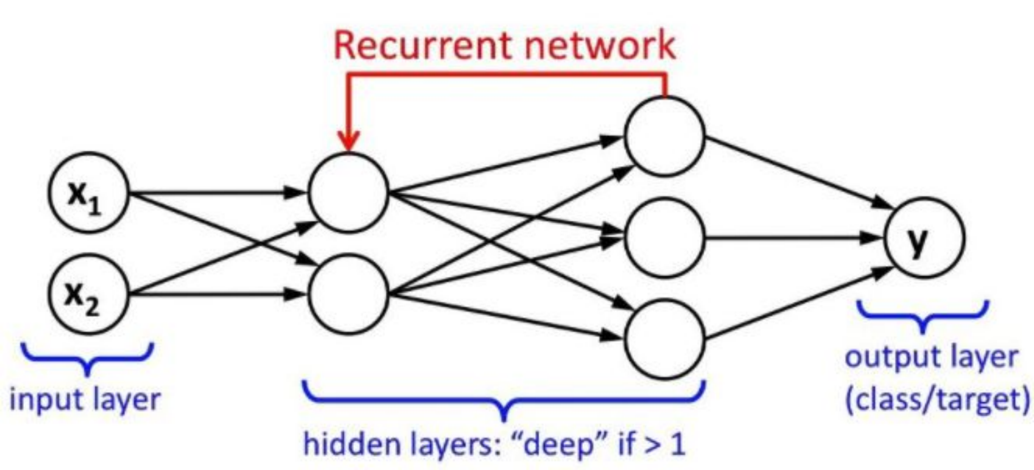


Figure 14: Image of RNN [7]

- Recurrent Neural Networks (RNN)
  - Allow us to read multiple images

DEEP LEARNING: NEURAL NETWORKS

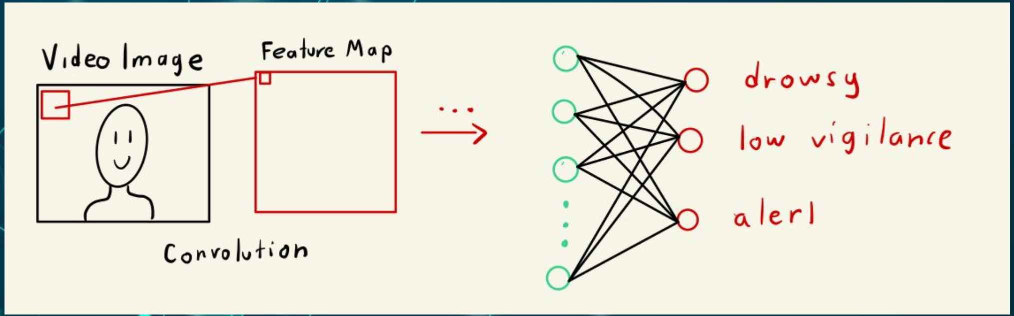


Figure 15: Image of CNN

- Convolutional Recurrent Neural Networks (CRNN)

DEEP LEARNING: NEURAL NETWORKS

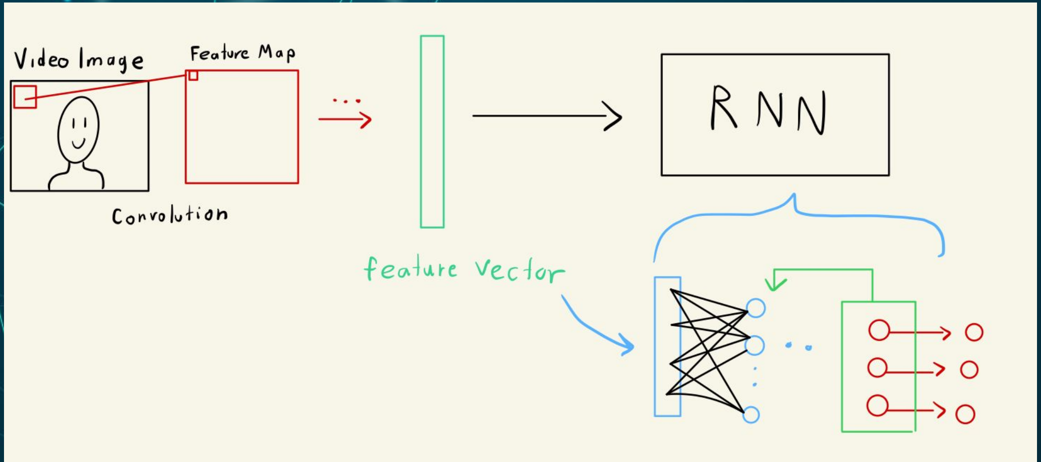


Figure 16: Image of CRNN

- Convolutional Recurrent Neural Networks (CRNN)