A No-Reference Autoencoder Video Quality Metric

Helard B. Martinez[1], Andrew Hines[2], and Mylène C.Q. Farias[1]
[1]University of Brasília, Brazil
[2]University College Dublin, Ireland
http://www.ene.unb.br/mylene

ICIP, 24 September 2019, Taipei, Taiwan

Motivation and Goals

Motivation and Goals

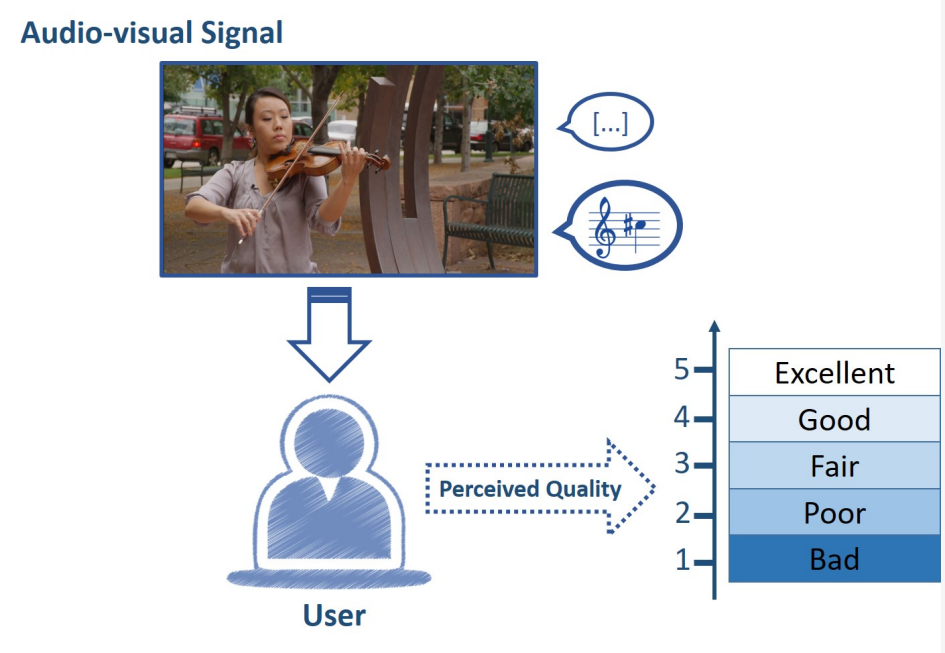
- Increase of types of MM services
- Quality of Experience (QoE) is an important aspect
- **Tools to quantify the quality of MM experience**

Motivation and Goals

Design a NR pixel-based video quality metric

- Created a large audio-visual dataset (diverse content and degradations);
- Collected video quality responses using an immersive methodology;
- EXtracted spatial-temporal features from the videos;
- Used Autoencoders algorithms to produce select the 'best' visual features;
- Mapped these visual features into subjective quality scores.

Subjective Experiments



Immersive Methodology (M. Pinson)

- Increase content diversity;
- Use longer videos, which convey an idea, with their audio;
- Keep the experiment interesting and reduce fatigue.

Apparatus and Physical Conditions

- Experiments divided into 3 sessions: Display, Training, Main;
- Scores collected (ACR scale, 5 points): MQS_{HRC} - Mean Quality Score (HRC)
- Recording Studio: @University of Brasilia
- Desktop computer, LCD monitor, set of earphones, Sound card Asus Xonar DGX 5.1
- Viewing conditions: ITU Rec. BT.500

Dataset

- Distortions: Bitrate compression, Packet-Loss, and Frame-Freezing;
  - **Video coding:** H.264 and H.265 video codecs (200 to 16,000 kbs);
  - **Packet Loss:** loss rates 0.01 to 0.10
  - **Frame freezing:** pauses (# pauses 1 to 3 - Length 1s to 3s)
- HRC: Hypothetical Reference Circuit
- Packet-loss and frame-freezing did not happen simultaneously

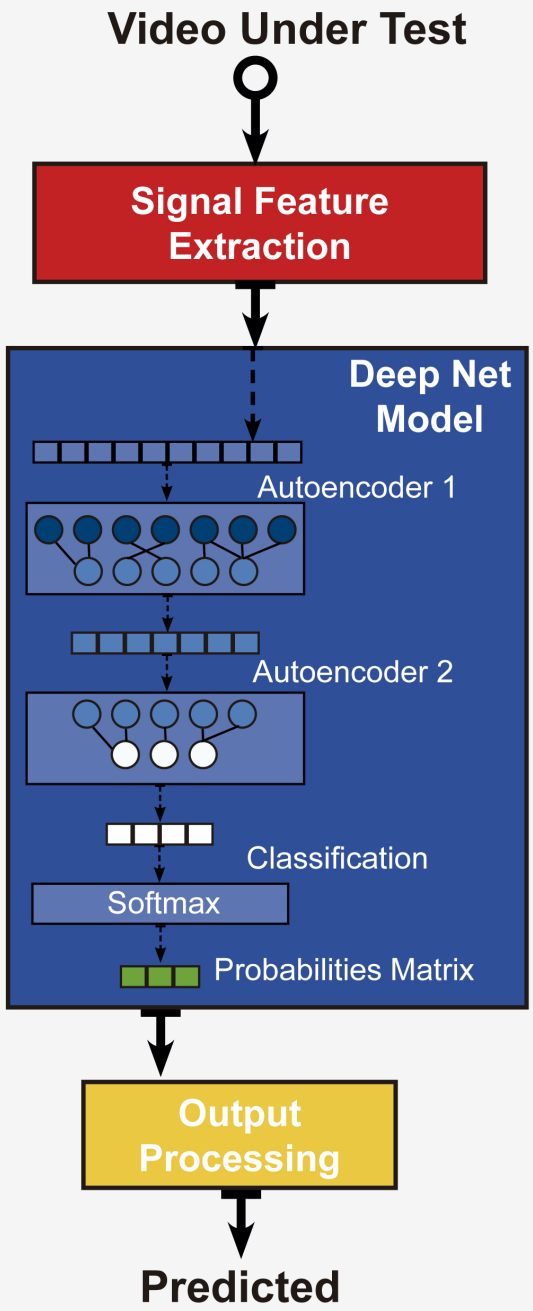
Dataset: Experiment 1

	Subjective Experiment
Participants:	60
Age Range:	19 - 36
Source Stimuli:	**60**
Test Conditions (HRC):	**12**
Test Sequences:	**720**
Average length:	37 sec
Temporal resolution:	1280x720 (720p)
Spatial resolution:	30 fps
Color space format:	4:2:0
Bit-depth:	16 bits
Sample frequency:	48 kHz
Audio codec:	PCM

Deep Autoencoder Model

- Deep Autoencoders: **select a visual features set with lower dimension and good description capacity;**
- Training (k-fold):
  - **Global Feature Matrix**: dataset spatial and temporal features
  - **Global Target Matrix**: dataset quality scores target
- NR Video Quality Metric.

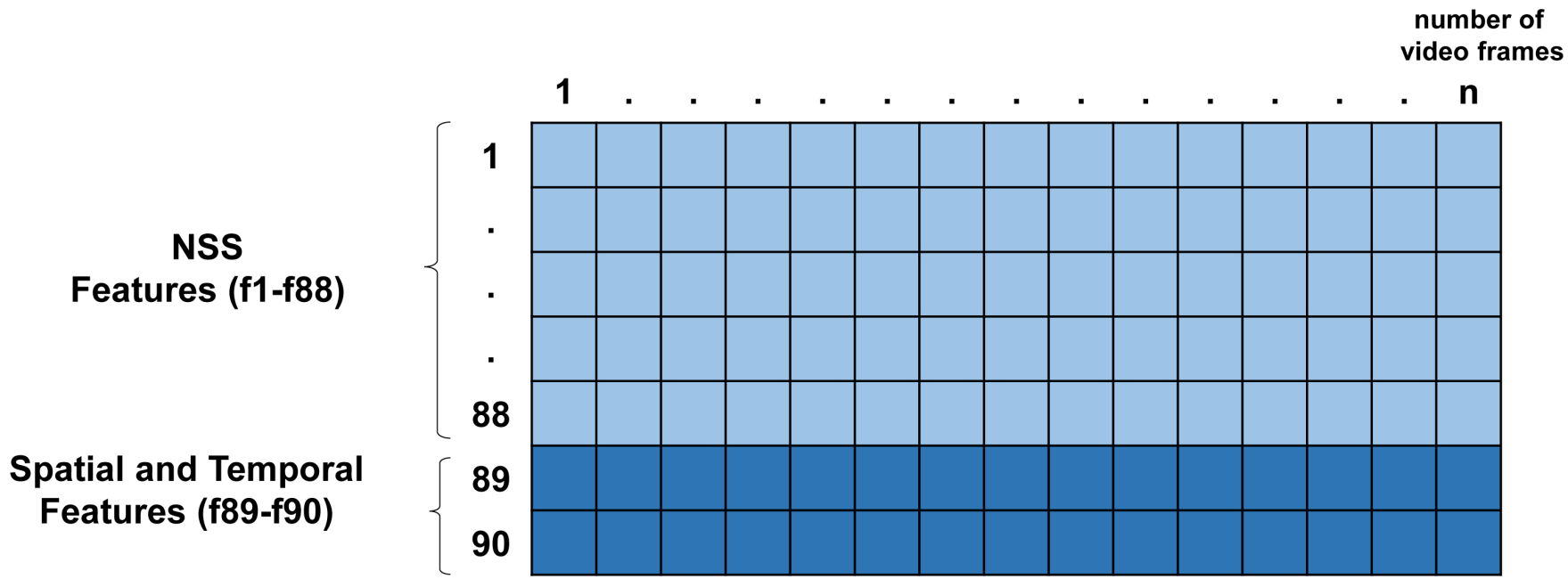
Quality Metric Diagram



Feature Extraction

Visual Features (90 features)

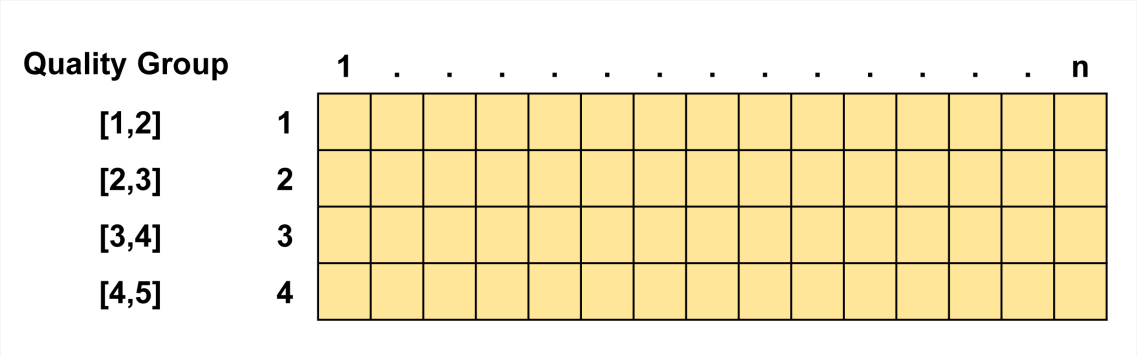
- (88) Natural Scene Statistics features - (2014, Zhang)
- (2) Spatial and Temporal features - (2007, Ostaszewska)
- 90-by-n matrix (n: number of frames)

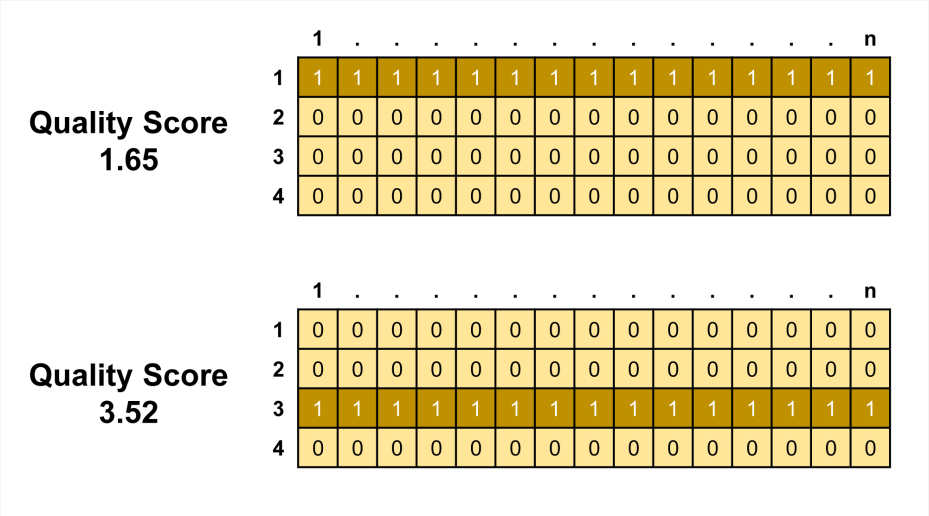


Model Training

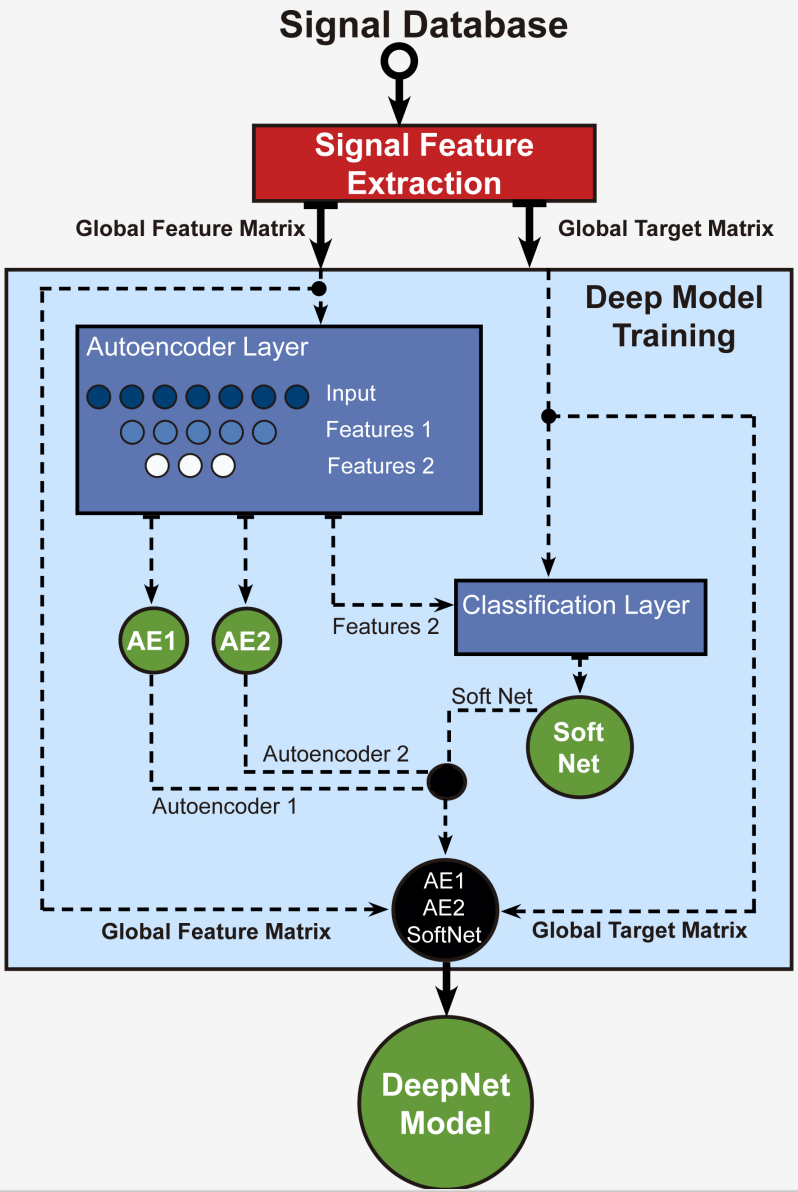
Target Set

- Built using subjective scores, for training
- 4-by-n matrix: 4 quality groups (ACR) and n video frames





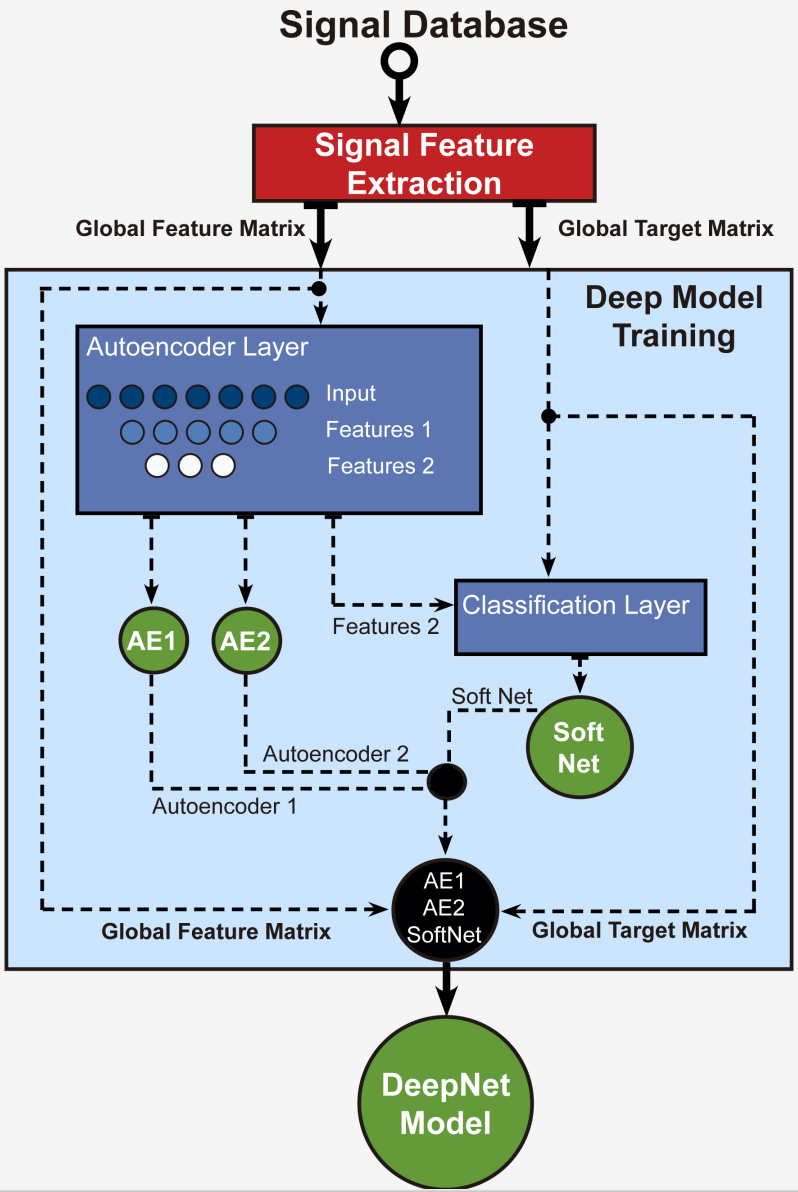
Model Training



Autoencoder Layer

- Input:
  - Global feature matrix
- Two autoencoders:
  - **Features 1, AE1**
  - **Features 2, AE2**

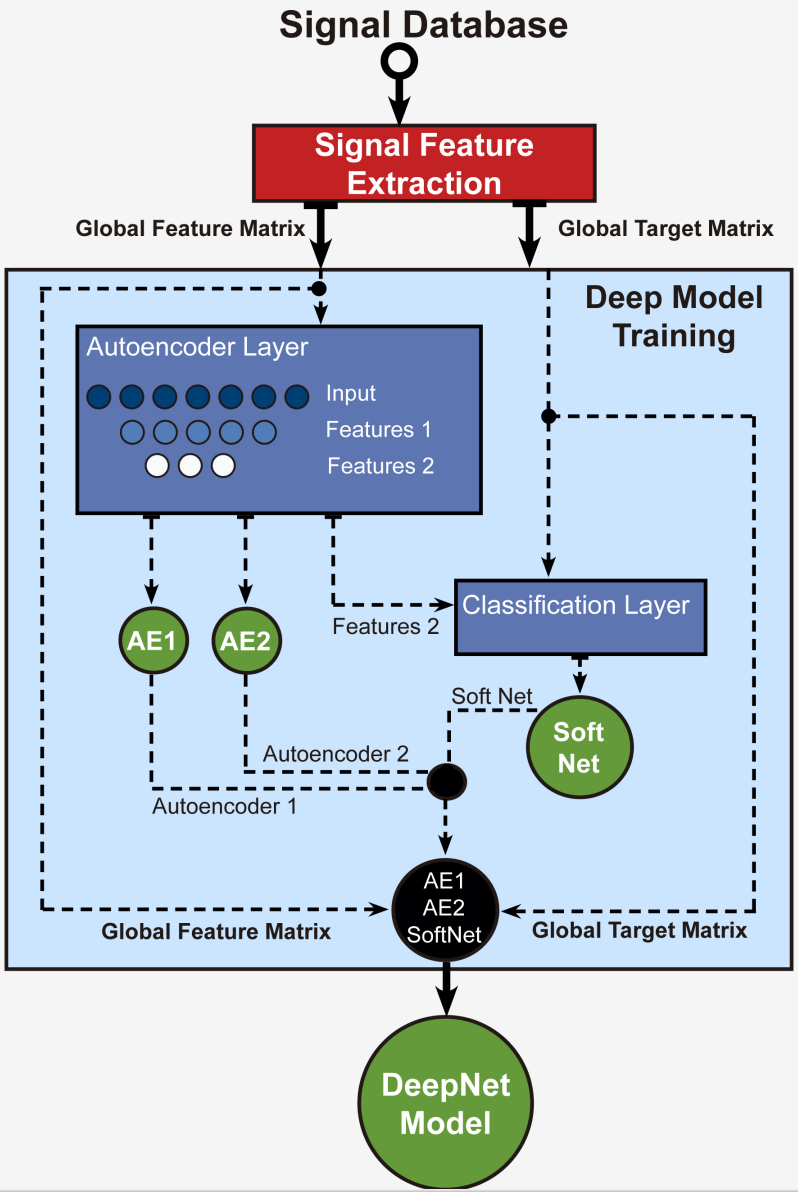
Model Training



Classification Layer

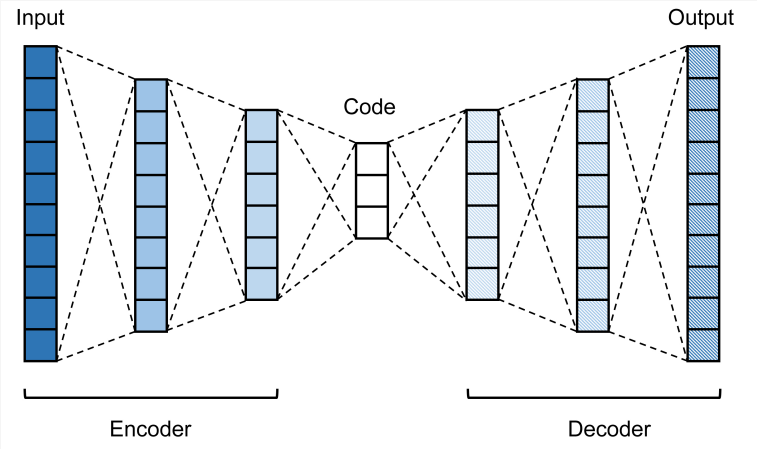
- Input:
  - Features 2
  - Global target matrix
- Softmax layer
  - **Soft Net**

Model Training

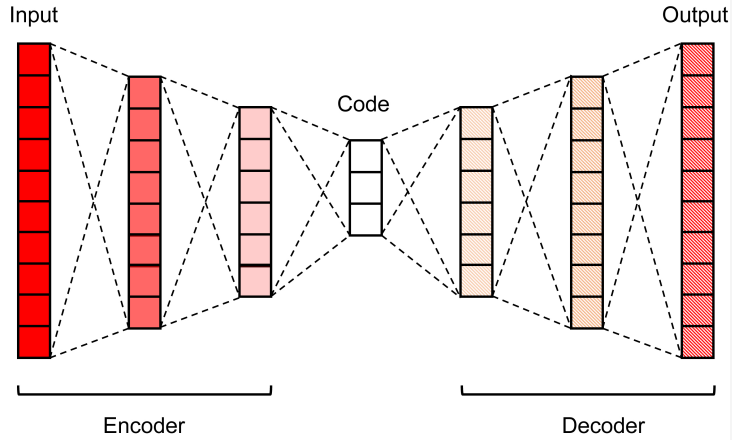


Deep Network Model

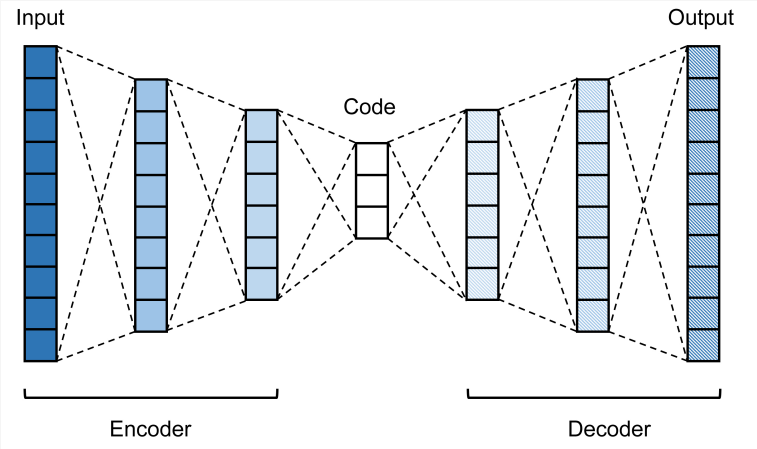
- Encoder1 (AE1)
  + Encoder2 (AE2)
  + SoftNet
- Trained with:
  - Global feature matrix
  - Global target matrix
- **DeepNet Model**



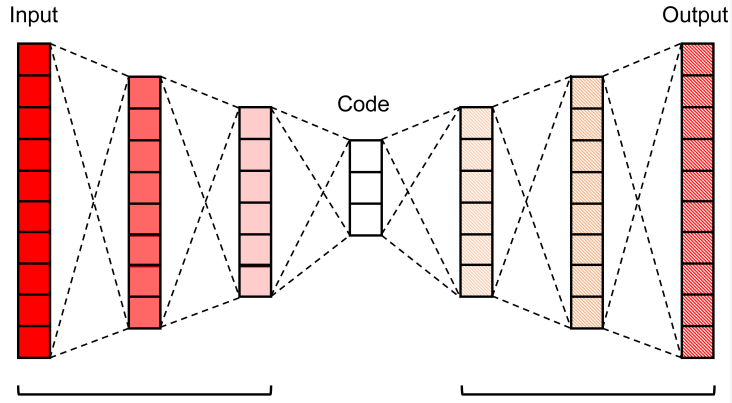
Autoencoder 1



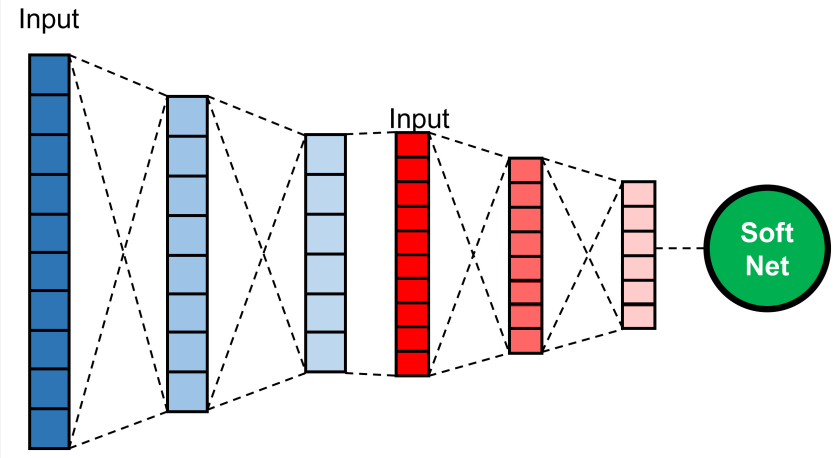
Autoencoder 2



Autoencoder 1



Autoencoder 2



DeepNet Model

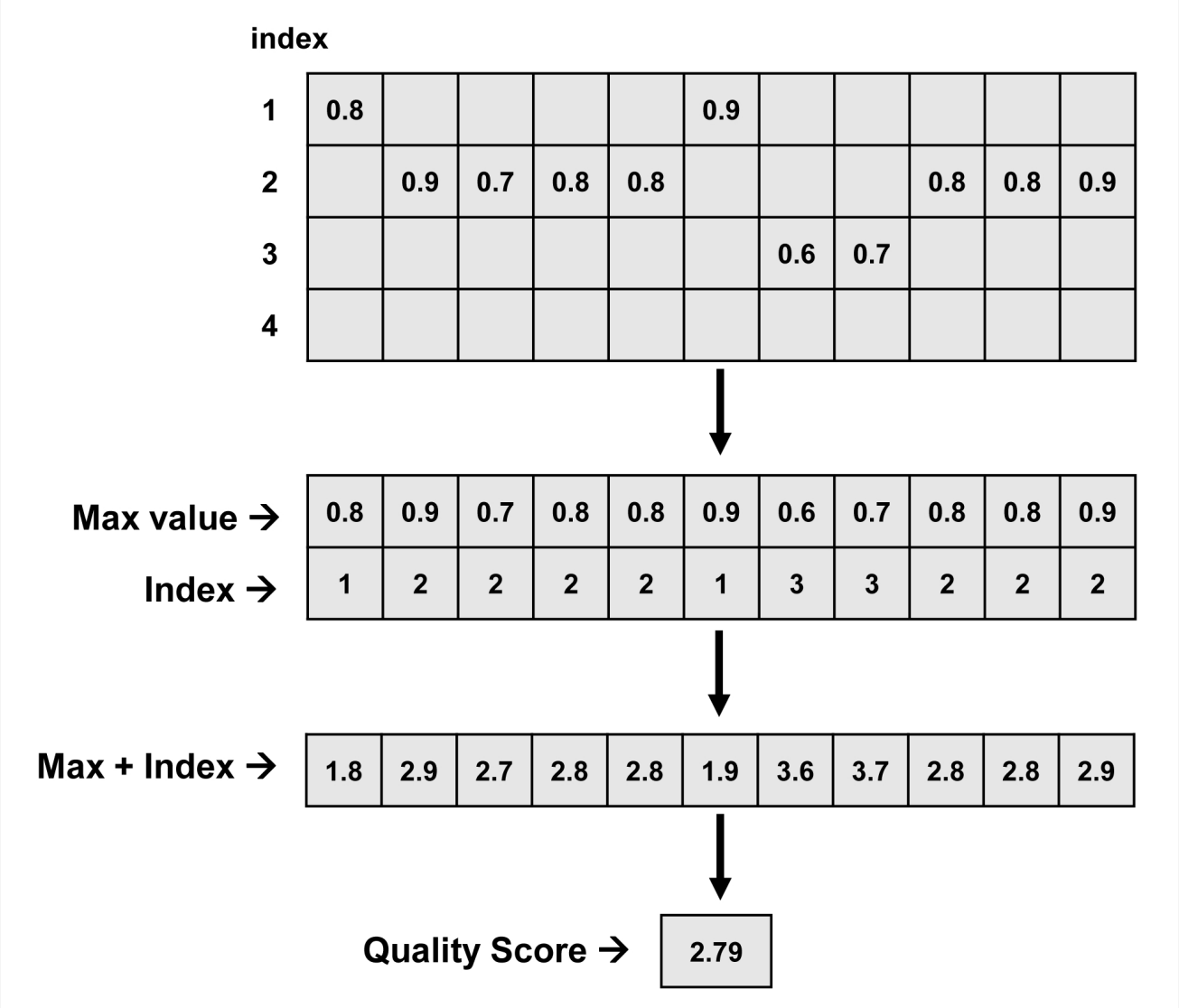
Model Training

Training Parameters

Layer	Parameters	Audiovisual Model
Autoencoder Layer	**Input**	90-by-N matrix
	Layer size #1	50
	Layer size #2	20
	Decoder transfer function	Linear
	L2 weigth regularization	0.001
	Sparsity Regularization	4
	Sparsity Proportion	0.05
Classification Layer	**Input**	20-by-N matrix
		4-by-N matrix
	Loss Function	Cross Entropy
Additional Info	**Training Set**	Exp. 1 of Dataset
	# sequences	720
	Method	10-fold CV

Model Performance

Output Processing

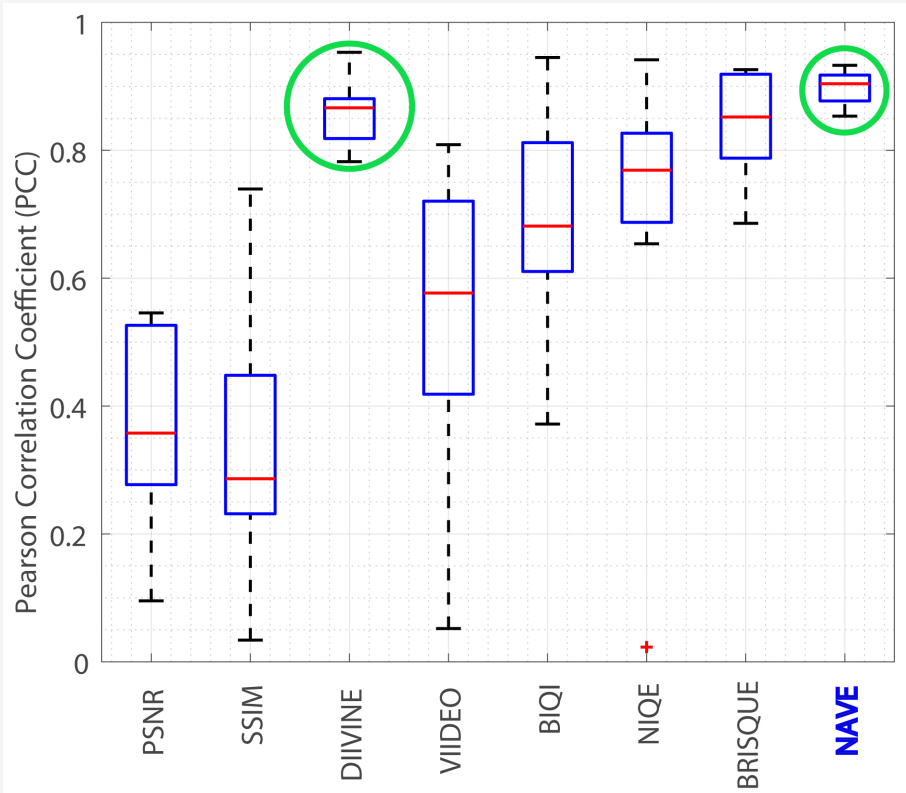


Model Performance

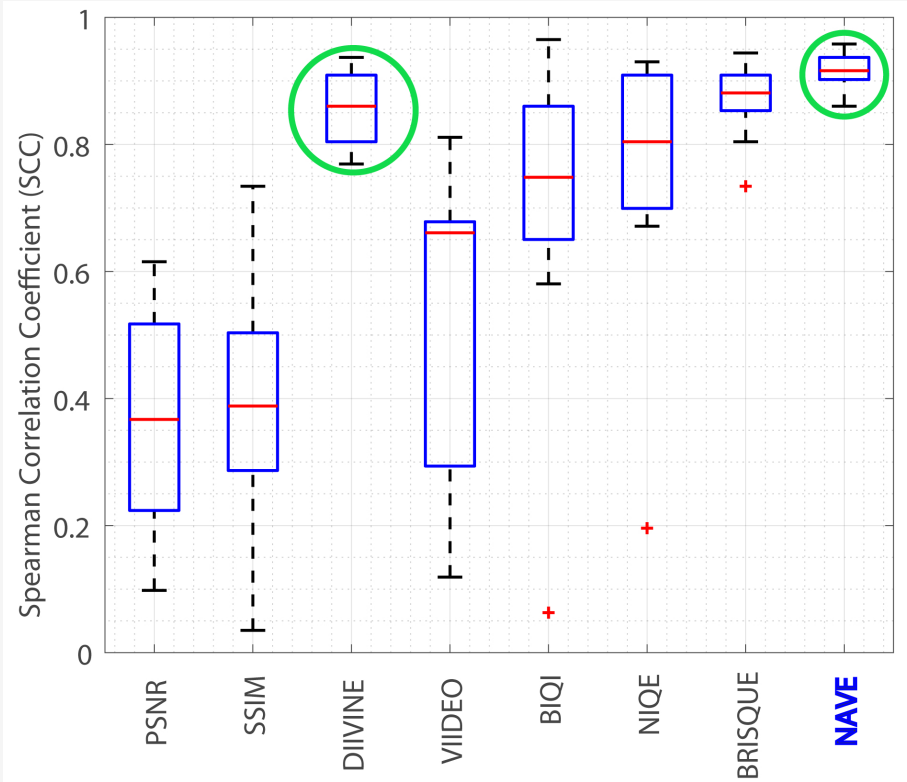
Metrics for comparison

- Video:
  - FR: SSIM, PSNR
  - NR: DIIVINE, VIIDEO, BIQI, NIQE, BRISQUE

NR Video Quality Metric



(PCC)



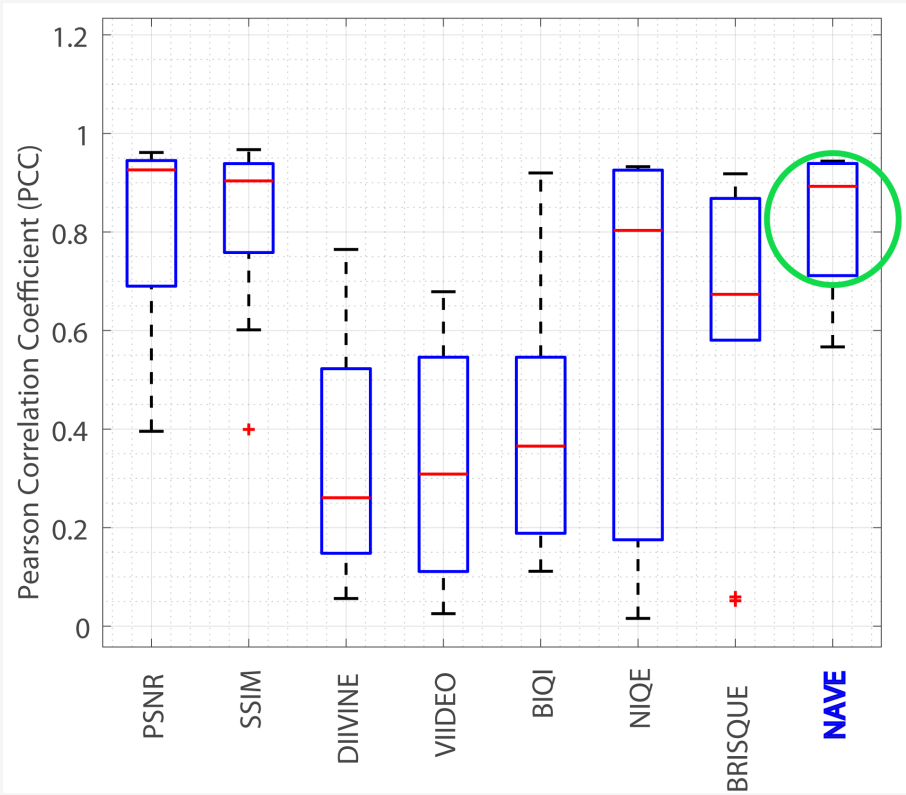
(SCC)

Model Performance

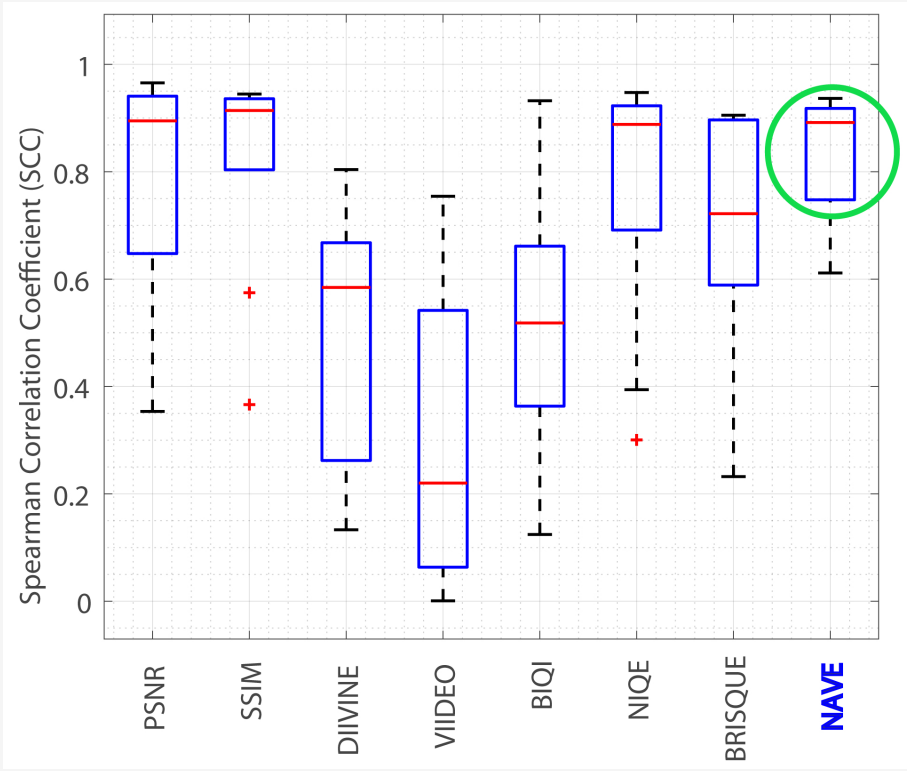
External Database: LiveNetflix-II

- 420 sequences AV Full-HD
- 15 source sequences, 7 network conditions, 4 bitrate adaptation strategies

NR Video Quality Metric - LiveNetflix-II



(PCC)



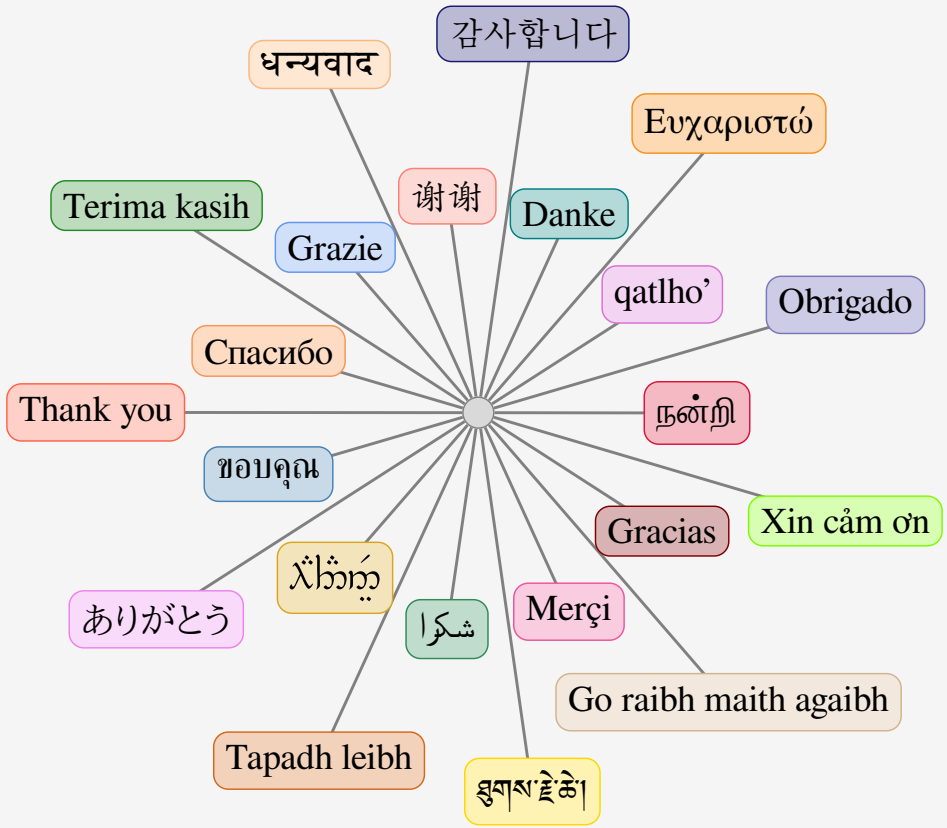
(SCC)

Conclusions:

- Used visual features to design a NR-VQ;
- Trained on a diverse content dataset;
  - Compression and transmission degradations;
  - Immersive subjective experiment;
- A Deep Autoencoder Model for quality assessment;
- Performed well, when compared to state-of-the-art metrics;

Future Work:

- Refine training parameters;
- Test additional descriptive features;
- Training on different databases.



Questions?

mylene@ene.unb.br,

http://www.ene.unb.br/mylene

http://www.ene.unb.br/mylene/databases.html