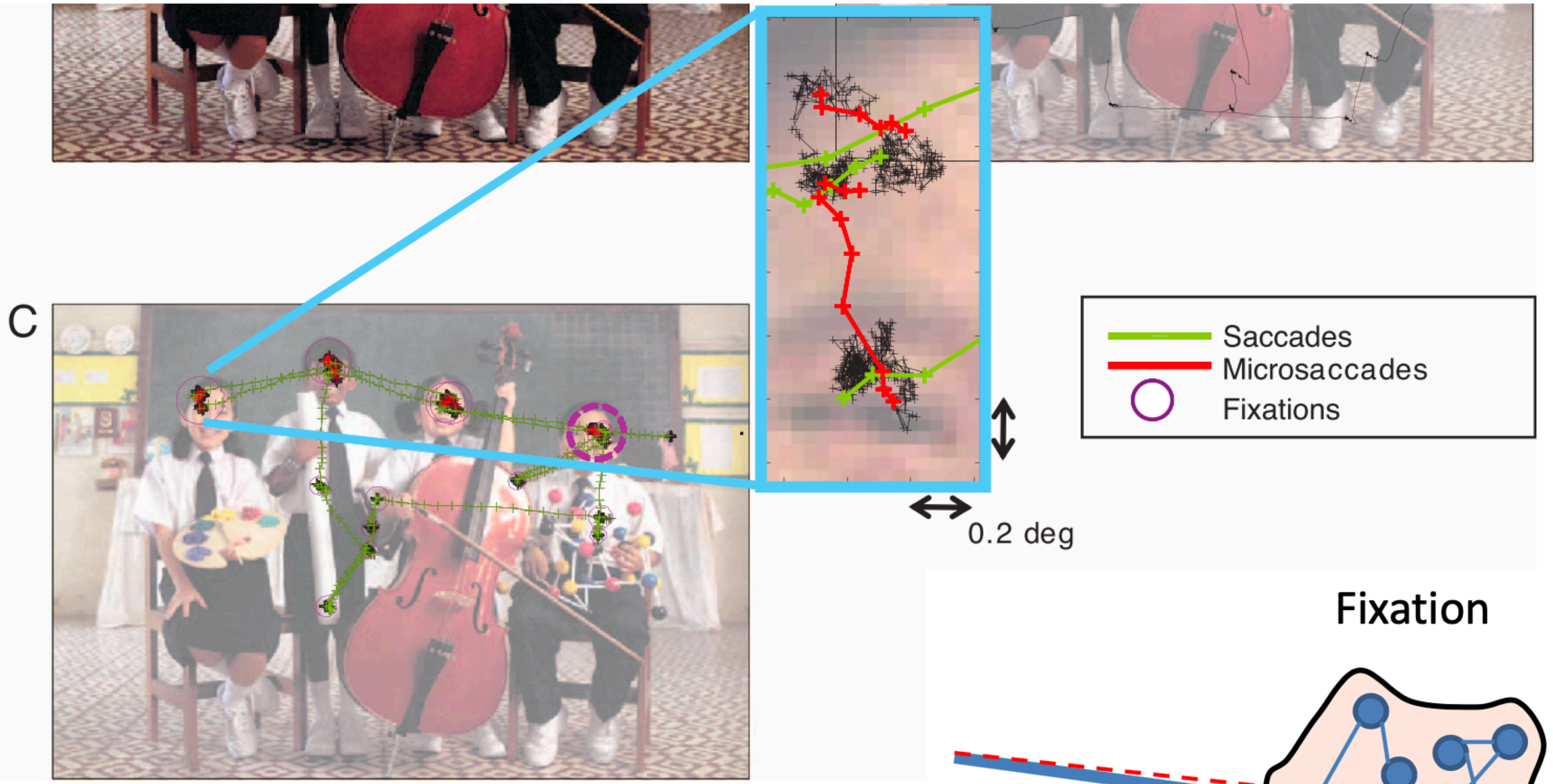
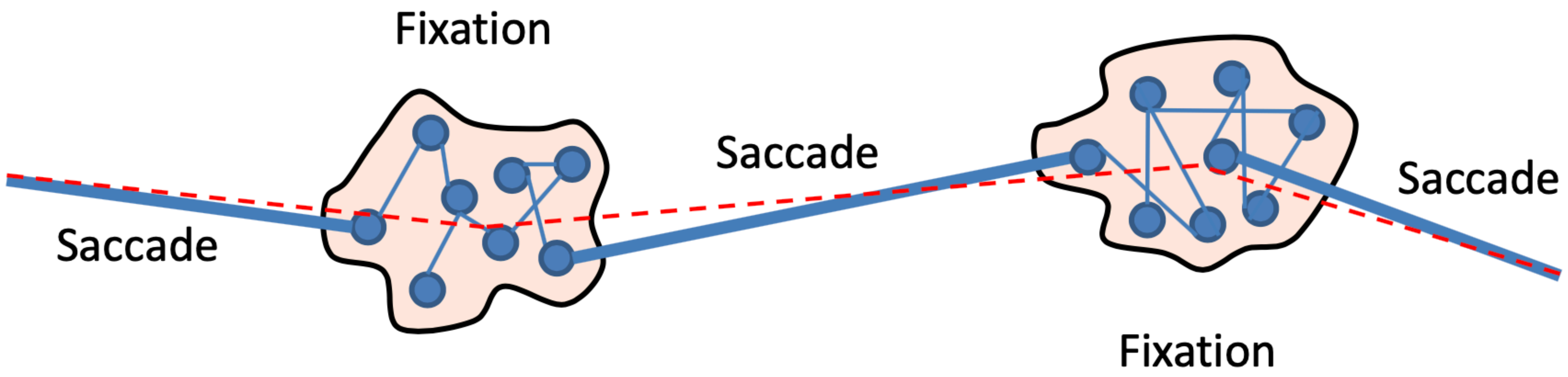
GazeMAE:
General Representations of **Eye Movements** using a Micro-Macro Autoencoder

Louise Gillian C. Bautista, Prospero C. Naval, Jr.
lcbautista1@up.edu.ph, pcnaval@dcs.upd.edu.ph
University of the Philippines

25th International Conference on Pattern Recognition (ICPR)

Introduction





Eye movements reveal a lot about our **perception, thought, and decision-making processes**.

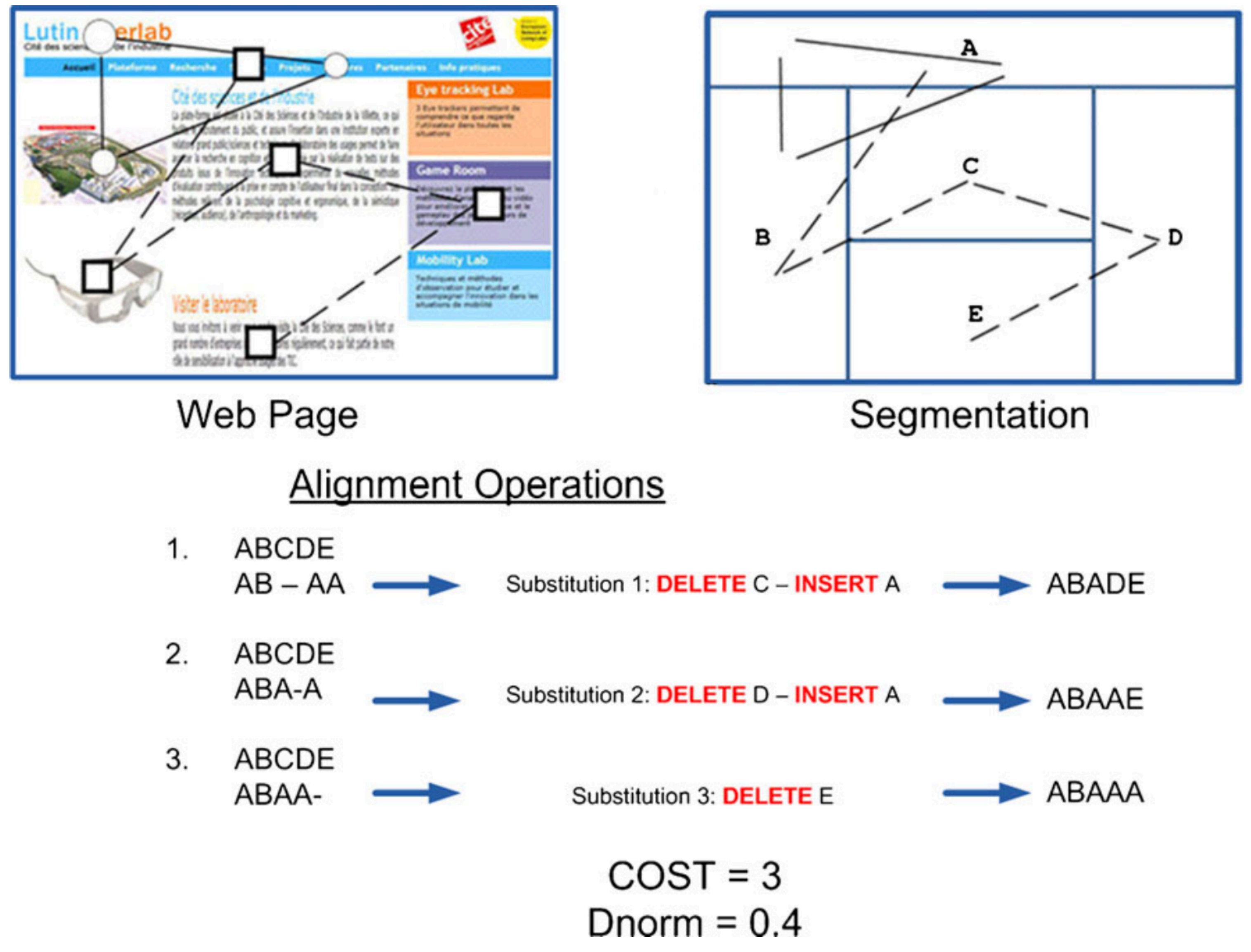
The amount of time we fixate, the speed at which we shift our gaze and even the almost invisible micromovements all hint at our cognitive processing.

They are valuable but **non-intrusive biosignals** that are useful for real-world applications such as education, healthcare, and security.

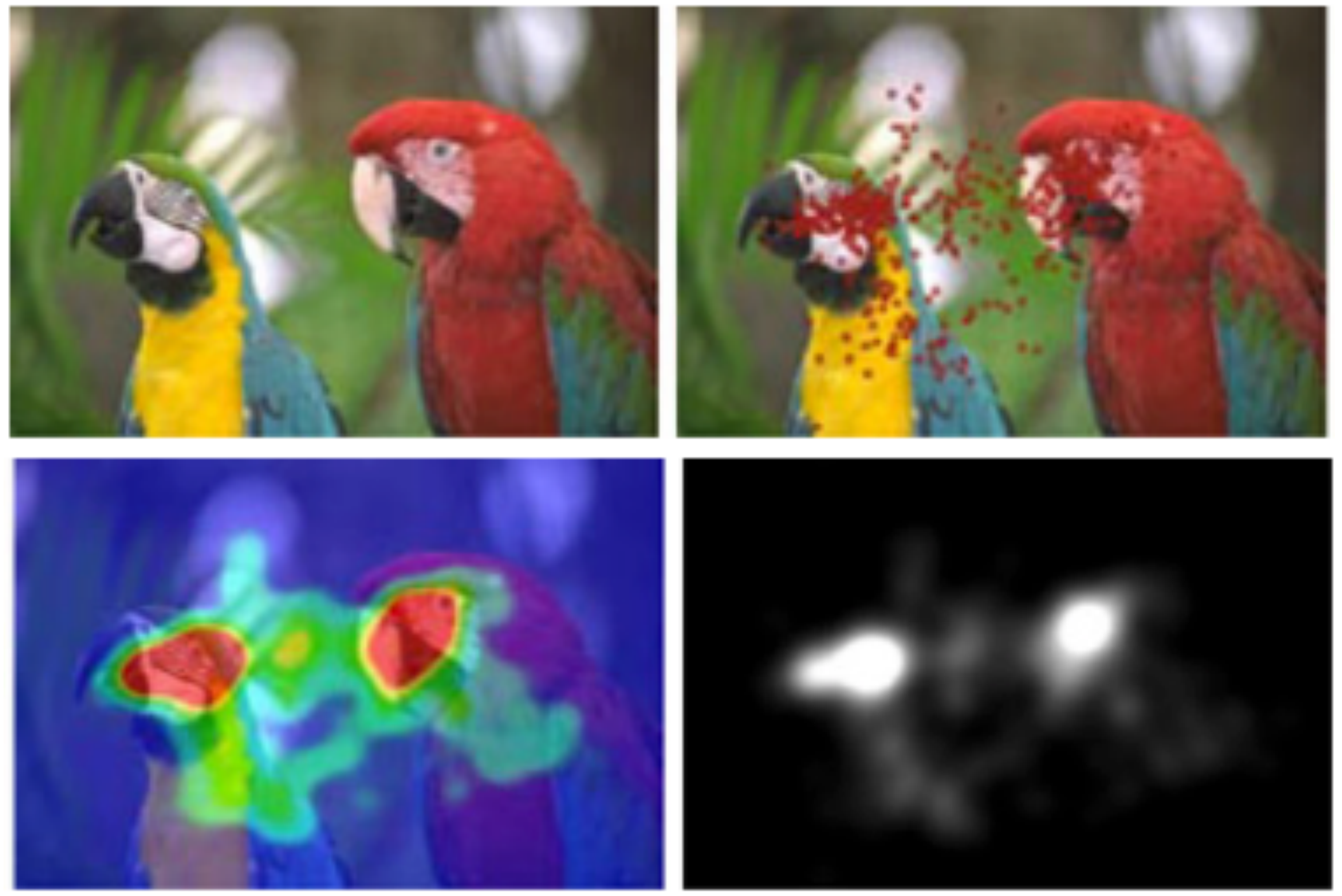
Introduction

However, classic computational methods to study and represent eye movements **cannot exploit the dynamic nature of eye movements as a result of aggregation and feature engineering**.

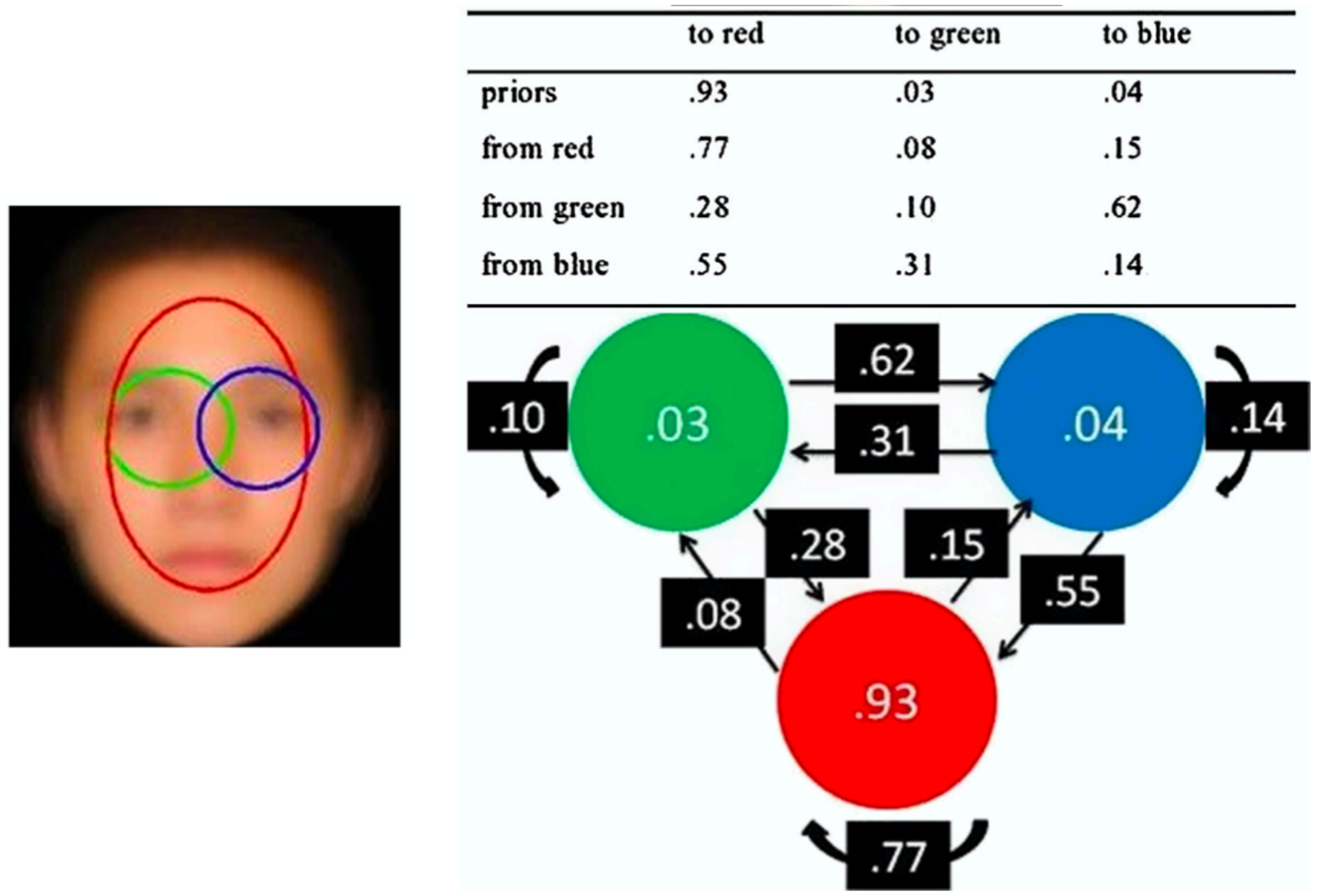
They may also be **stimuli-dependent**, placing the restriction that eye movements have to come from the same stimuli.



Methods for comparing scanpaths and saliency maps: strengths and weaknesses (Le Meur and Baccino, 2013)



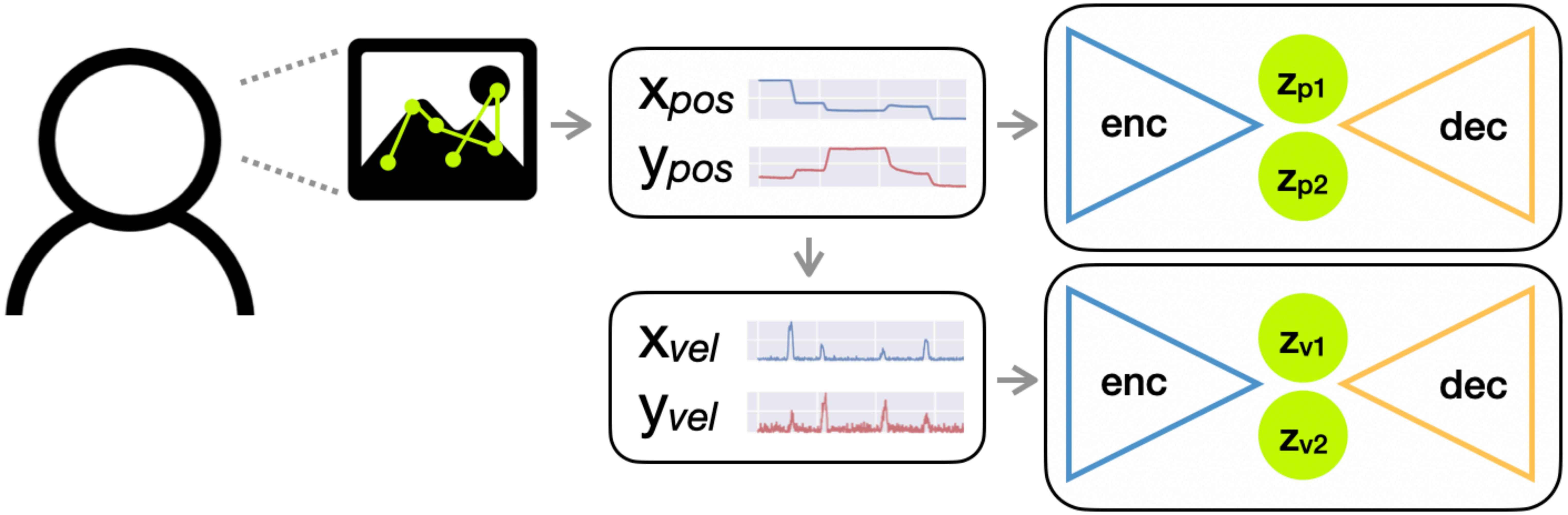
Methods for comparing scanpaths and saliency maps: strengths and weaknesses (Le Meur and Baccino, 2013)



Hidden Markov model analysis reveals the advantage of analytic eye movement patterns in face recognition across cultures (Chuk et al., 2017)

Methodology

Learn an abstract representation of eye movements that highlight and preserve both micro and macro movements — by using an **autoencoder with dilated temporal convolutional networks**



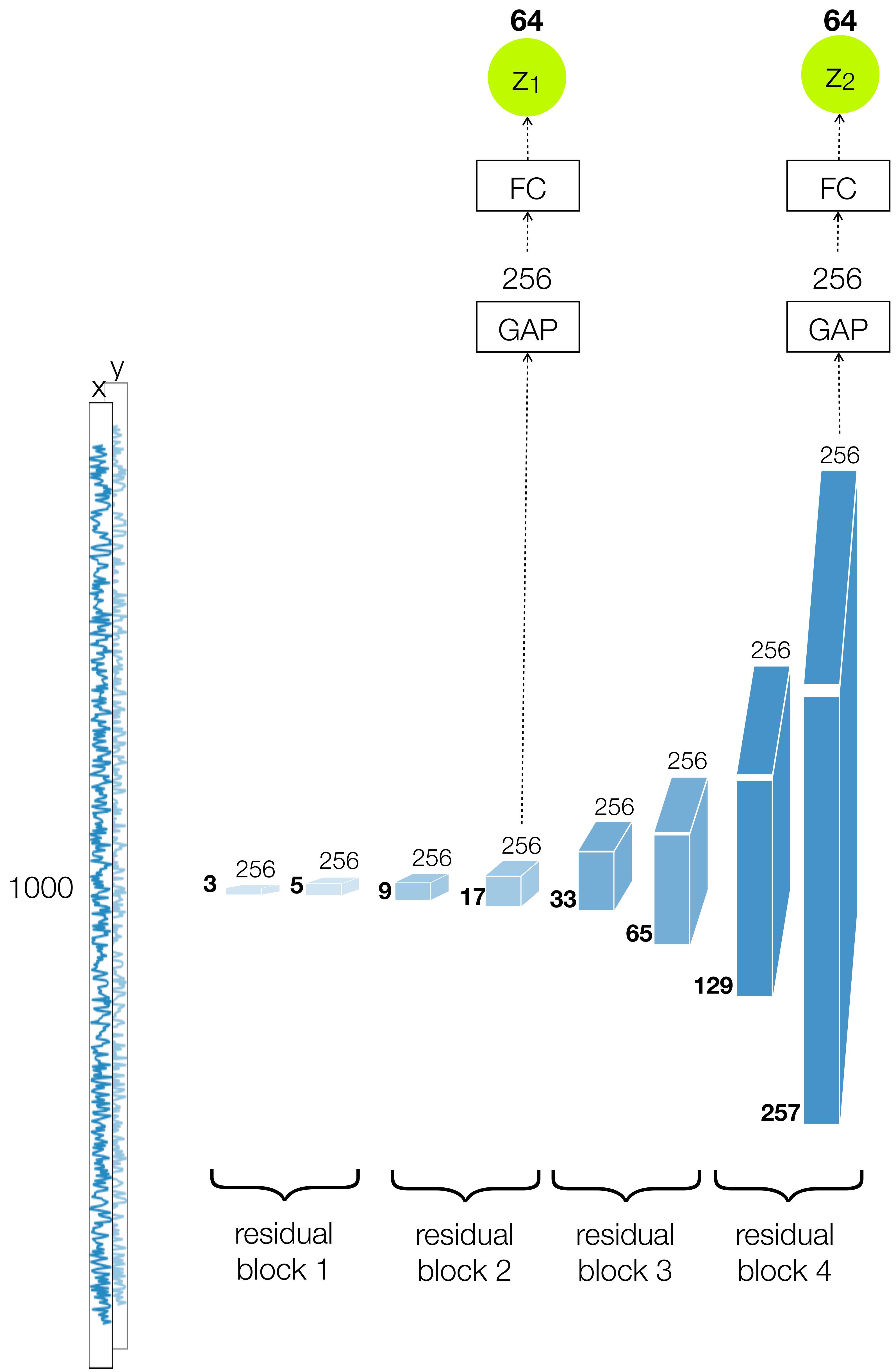
Raw eye movement data are treated as signals
(position and velocity)

Methodology

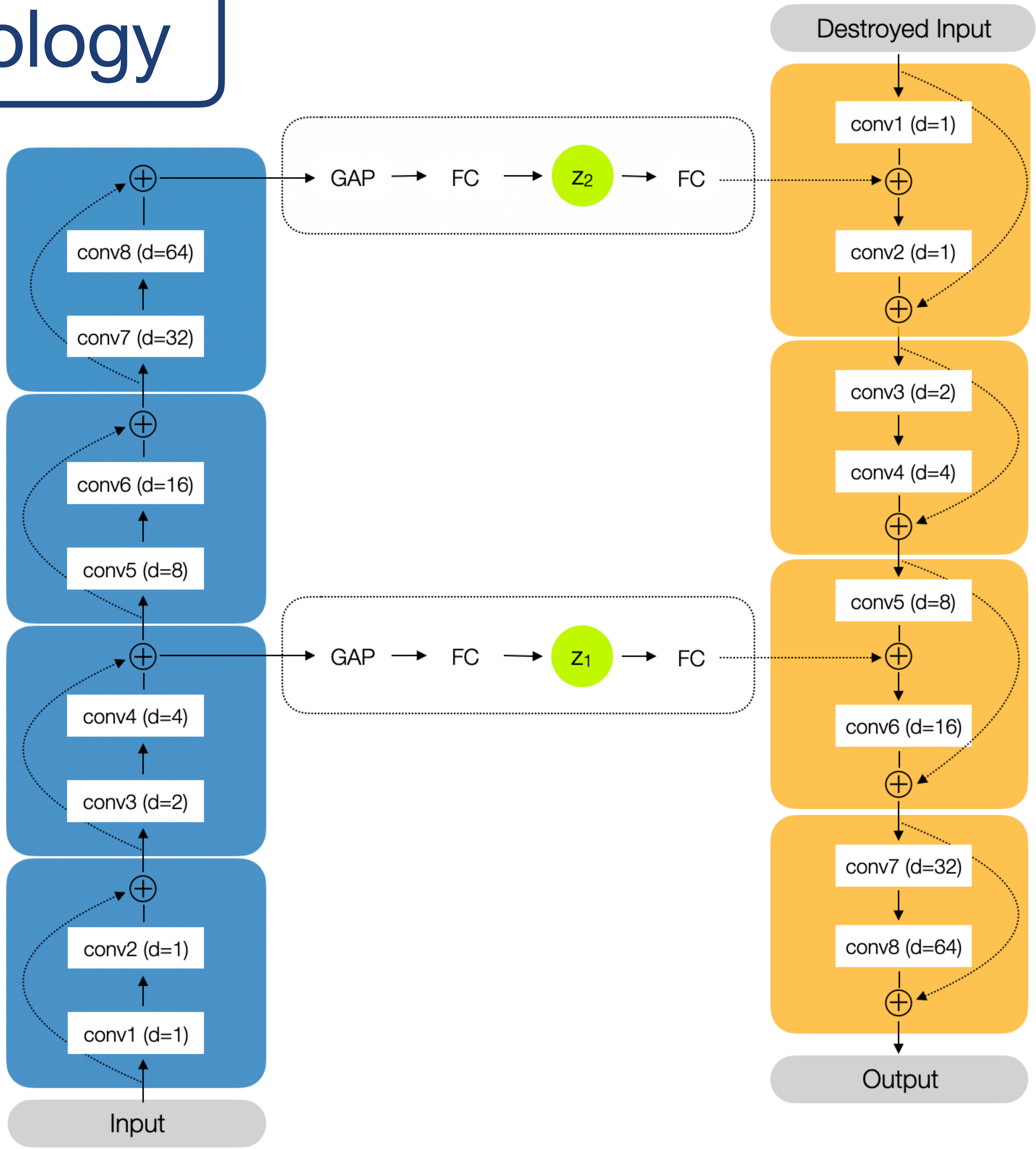
In **dilated TCNs,**

The receptive field grows exponentially across layers, which means the **encoded context and information is also different at each layer.**

The AE has two bottlenecks. The one at the **fourth layer corresponds to the micro representations, while the one at the eighth layer is for the macro.**



Methodology



Encoder and decoder both have 4 convolution blocks.

The decoder is interpolative: **we feed a destroyed version of the input to the decoder** to reconstruct, instead of having it predict one value at a time (i.e. autoregressive).

This gave the same performance but at less training time.

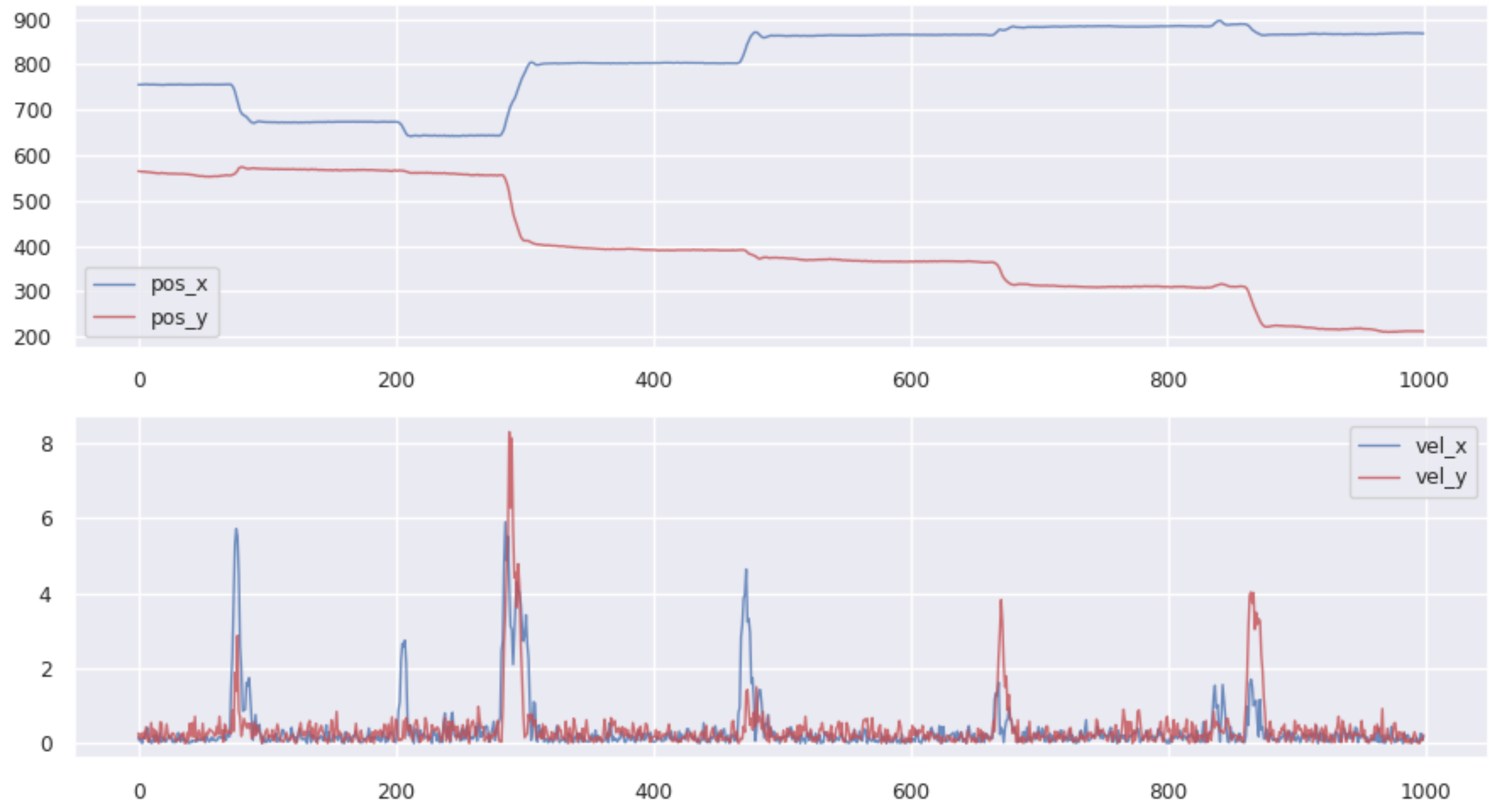
Data

	Hz	Stimuli	Tasks	Subj.	Sample	Time(s)
EMVIC	1000	face	free	34	1430	ave. 2.5s
FIFA	1000	natural	free, search	8	3200	2s
ETRA	500	natural, puzzle	free, search	8	480	45s
Total				50	5110	

Augmented by taking overlapping 2s windows.
Total samples after augmentation: **68,178**

1000 Hz data sets are downsampled to 500 Hz.

Example of position and velocity signals



Training

Network Parameters

	position AE (AE_p)	**velocity AE** (AE_v)
Encoder TCN	128 filters x 8 layers	256 filters x 8 layers
Micro-scale Bottleneck	64-dim FC	64-dim FC
Macro-scale Bottleneck	64-dim FC	64-dim FC
Decoder TCN	128 filters x 4 layers; 64 filters x 4 layers	128 x 8 layers
Total Parameters	652,228	1,964,676

Learning Rate: 5e-4
Optimizer: Adam
Batch Size: 256 (pos), 128 (vel)
Epochs: 14 (pos), 25 (vel)

Framework: PyTorch
GPU: GTX 1070

Afterwards, the representations will be evaluated on **classification tasks with a linear SVM.**

Results

Representations outperform previous works

Velocity is important for eye movement biometrics, position is important for inferring the stimuli

Classification Task	PCA_{pv}	z_p	z_v	z_{pv}	others
Biometrics (EMVIC-Train)	18.4	31.8	**86.8**	84.4	86.0 [26]
Biometrics (EMVIC-Test)	19.7	31.1	**87.8**	87.8	81.5 [26] 82.3* 86.4*
Biometrics (All)	24.6	29.0	79.8	78.4	-
Stimuli (4)	38.8	81.3	85.4	87.5	-
Stimuli (3)	55.8	90.3	87.2	**93.9**	88.0** [29]
Age Group	62.0	61.9	77.7	77.3	-
Gender	51.12	54.9	85.8	86.3	-

Also robust against viewing time

Can handle 1s of data to up to 45s without loss of performance

Classification Task	1s	2s	2s*	full
Biometrics (EMVIC-Train)	78.9	84.2	83.35	86.8 (22s)
Biometrics (EMVIC-Test)	79.0	85.6	86.6	87.8 (22s)
Biometrics (All)	69.3	76.9	79.7	79.8 (45s)
Stimuli (4)	46.7	59.2	85.0	85.4 (45s)
Age Group	75.1	78.2	-	-
Gender	79.4	85.9	-	-

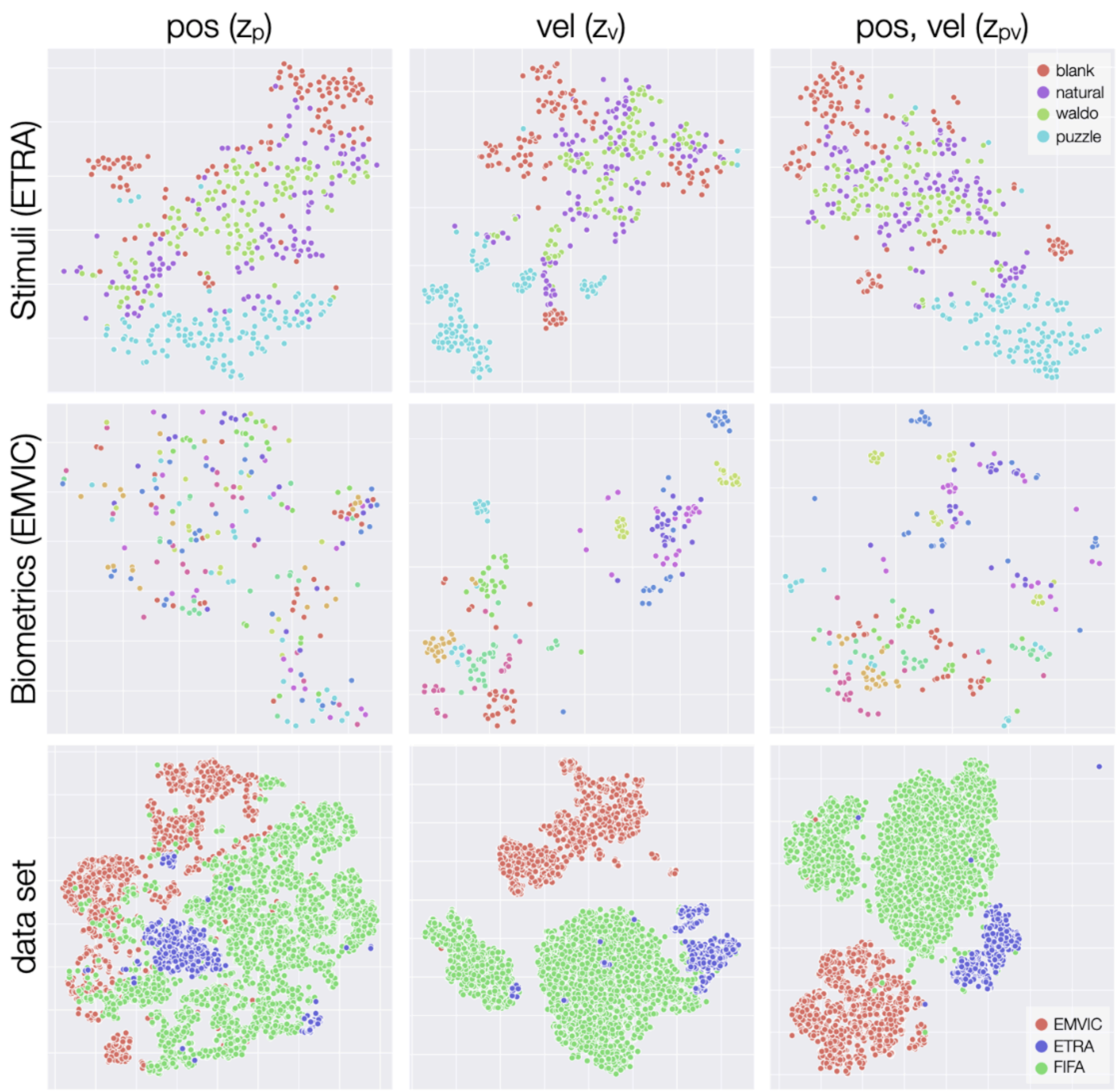
Model generalizes to an unseen dataset

And outperforms a model trained solely on that unseen dataset (MLR)

Classification Task	AE_v	AE_v-250	AE_v-MLR
Biometrics (MIT-LowRes)	23.7	21.5	18.38

Results

t-SNE Plots



Conclusion

This work proposed an autoencoder that learns learns **micro and macro-scale representations** for eye movements.
Models were trained on both position and velocity signals.

Competitive results were achieved despite using only a linear classifier.
The model is found to be robust to viewing time, and generalizes to unseen samples from a different data set.