Improving Performance and Interpretability in Recognizing Facial Action Units with Deep Neural Networks

Ciprian A. Corneanu

December 12, 2019

Director | Dr. Sergio Escalera
Dept. de Matemàtiques i Informatica
Universitat de Barcelona

Co-director | Dr. Meysam Madadi
Dept. de Matemàtiques i Informatica
Universitat de Barcelona

UNIVERSITAT DE BARCELONA

Părinților mei
To Lenka

Outline I

The Human Face

Figure: The human face is a rich source of information.

Biological Structure of The Human Face

Figure: The human face is composed by soft tissue attached to a bone structure.

Hard Tissue Variability

Figure: Skull shape and size vary with age.

Hard Tissue Variability

Figure: Skull shape and size vary with gender, race and identity.

Soft Tissue Variability

Figure: Soft tissue variability with: (a) age, (b) race.

Facial Muscles and The Expressive Face

Figure: We use facial muscles to eat, speak, move our eyes and produce a wide range of expressions that convey information in social contexts[1].

[1](Left) Sculptures by Franz Messerschmidt, 18th century artist. (Right) Paintings by Duarte Vitoria contemporary artist.

Facial Muscles and the Facial Action Units

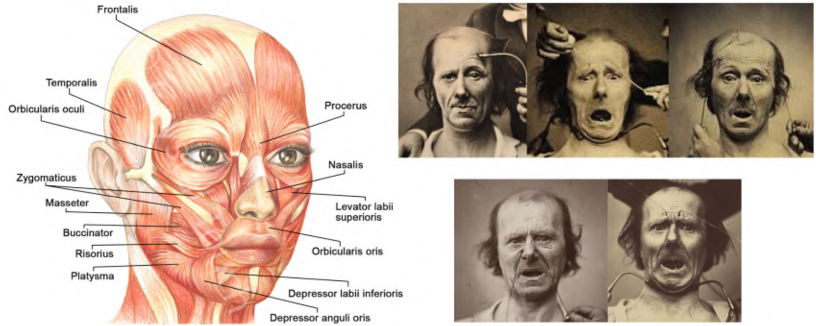


Figure: We can encode facial expressions based on muscular activity[2].

[2]Photographs from Duchenne de Boulogne, "Mécanisme de la physionomie humaine", (1862).

The Facial Action Coding System

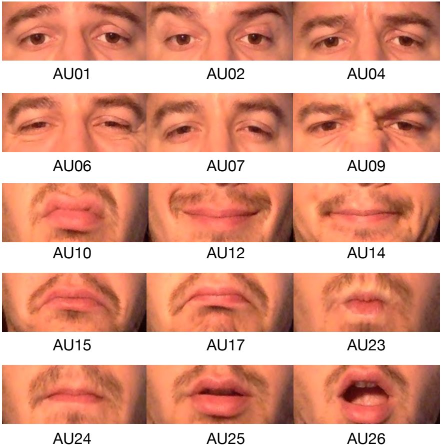


Figure: (a) The Facial Action Coding System[3] is a descriptive encoding of the expressive face. (b) An illustration of Action Units.

[3] Ekman, Paul and Wallace V. Friesen. "Facial action coding system: a technique for the measurement of facial movement." (1978).

Goal

The main goal of this thesis is to develop mathematical models that **learn to recognize Facial Action Units** with **high performance** by using **interpretable Deep Neural Networks**.

Outline I

Learning

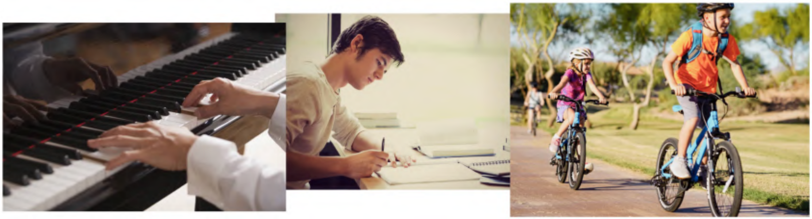


Figure: We associate learning with many different day-to-day experiences.

Two Ways for Acquiring Knowledge

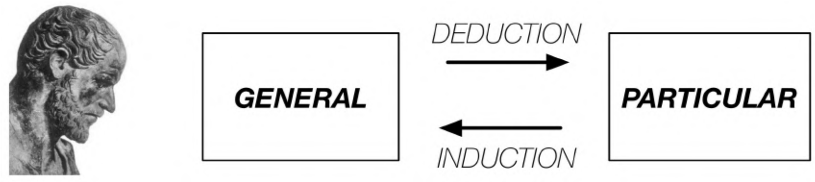


Figure: Aristotle was on of the first to state that there are two basic mechanisms for acquiring knowledge.

Formalizing Learning by Induction

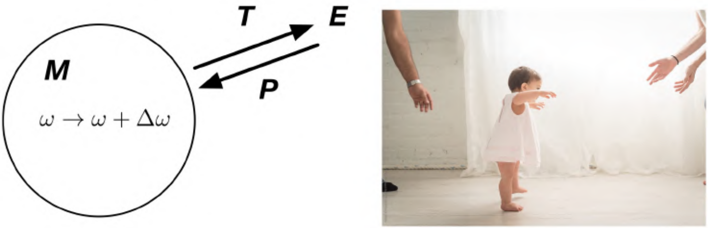


Figure: We associate learning with a variety of situations and experiences.

Definition. *A model M is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E*[4]*.*

[4] Mitchell, T. (1997). Machine Learning. McGraw Hill. p. 2.

The Problem of Generalization

For a model to generalize from a limited number of observations, the world needs to have:

- **Invariance.**[5] The context in which generalization is to be applied cannot be fundamentally different from that in which it was made.
- **Learnable Regularity.** Patterns can be recognized efficiently.

[5] Leslie Valiant, (2013), "Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World"

Measuring Generalization

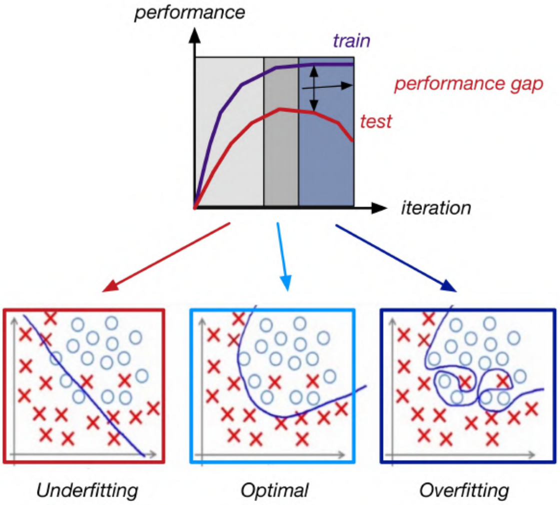


Figure: Ability to generalize is the fundamental property of learning models.

The "Tribes" of Machine Learning[6]

Figure: Machine Learning still lacks a unifying theory, instead, different paradigms persist.

[6] Domingos P. (2015). The Master Algorithm

Connectionists try to Reverse-engineer the Brain

Figure: A depiction of the human cortex. [7]

[7] Dunn G., Cortical Columns

Deep Neural Networks

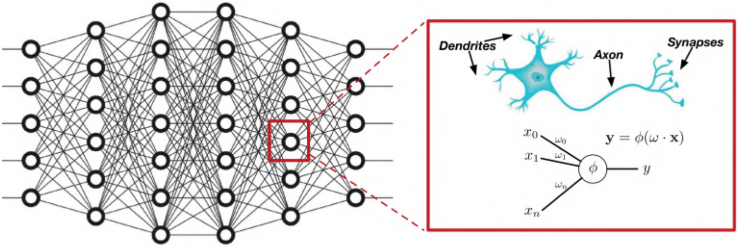


Figure: A Deep Neural Network (DNN) is an universal function approximator [8] and it can be optimized through back-propagation [9].

[8] Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators." Neural networks 2.5 (1989): 359-366.

[9] Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." Nature 323.6088 (1986): 533-536.

Convolutional Neural Networks and the New Rise of Connectionism

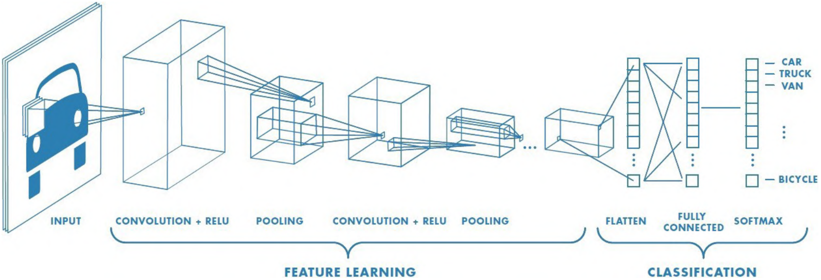


Figure: Convolutional Neural Networks are specially designed for learning representations from images.

Deep Neural Networks Today

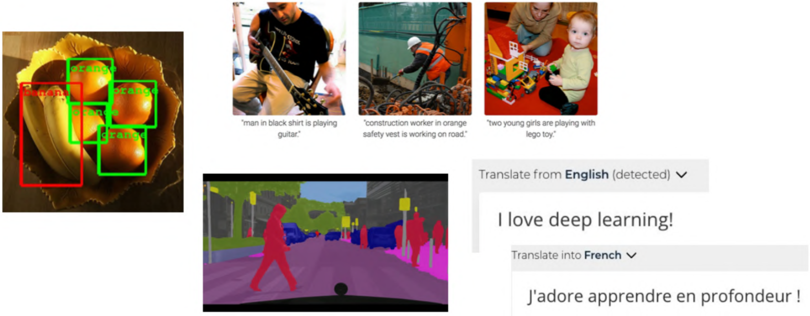


Figure: In recent years Deep Neural Networks (DNN) have become the state-of-the-art models in a myriad of tasks due their flexibility and high performance.

Performance, Complexity and Interpretability

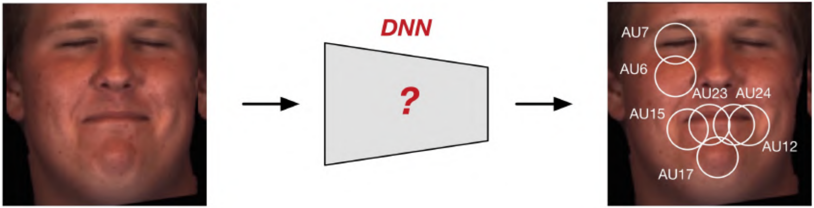


Figure: Despite flexibility and high performance deep neural networks are increasingly complex and opaque. They are regarded as “black-box” models. We simply don’t know when DNNs are learning, how are they learning and when will they fail.

Goal

The main goal of this thesis is to develop mathematical models that **learn to recognize Facial Action Units** with **high performance** by using **interpretable Deep Neural Networks**.

Goal

The main goal of this thesis is to develop mathematical models that **learn to recognize Facial Action Units** with **high performance** by using **interpretable Deep Neural Networks**.

Outline I

Overview

Figure: We learn to recognize Facial Action Units from images using Deep Neural Networks.

Characteristics of Action Units

(a) (b)

Figure: There are two fundamental characteristics of AUs. (a) AUs locally modify facial morphology. By masking just a small region an expressive face becomes indistinguishable from neutral. (b) AU recognition is a multi-label classification problem. Several AUs can be active at the same time and AU pairs can be strongly correlated.

Deep Structure Inference Network

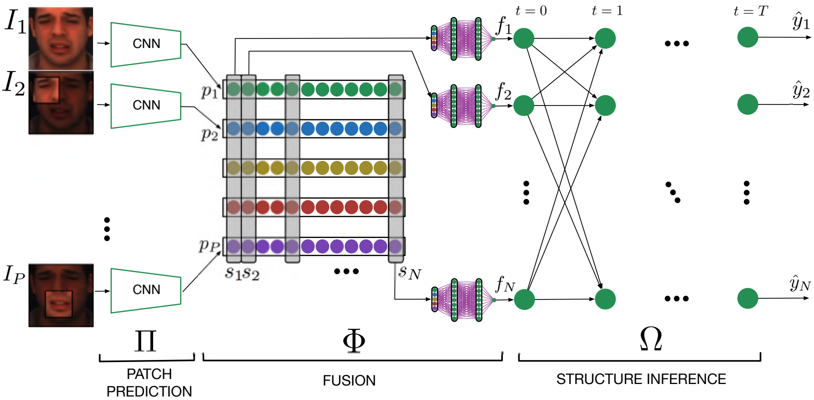


Figure: Deep Structure Inference Network (DSIN). DSIN learns independent AU predictions from global and local deeply learned features and replicates a message passing mechanism between AUs.

Deep Structure Inference Network: Details

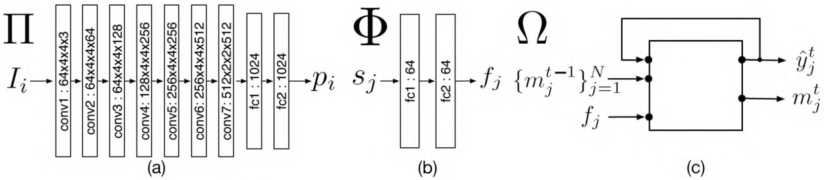


Figure: (a) Topology of patch prediction CNNs. Each convolutional block consists of a convolutional layer with stride 2 and batch normalization. The convolutional layer is shown by the number of filters followed by the size of the kernel. The last layers are fully-connected (FC) layers marked with the number of neurons. All neurons use ReLU activation functions. (b) Each fusion unit is a stack of 2 FC layers. (c) A structure inference unit. For better visualization, we just show the interface of the unit without the inner topology.

Message Passing for Structure Inference

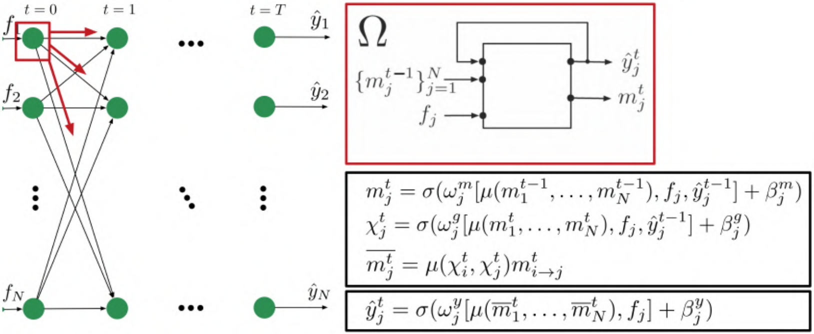


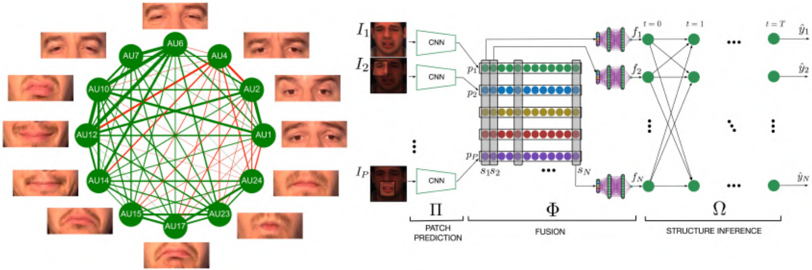
Figure: Structure inference in DSIN.

Datasets

	#seqs	#persons	#frames	#active frames	#AU	label cardinality†	label density‡
DISFA	27	27	130,814	56,356	10	3.04	4.05
BP4D	41	328	144,682	117,075	12	4.05	0.22

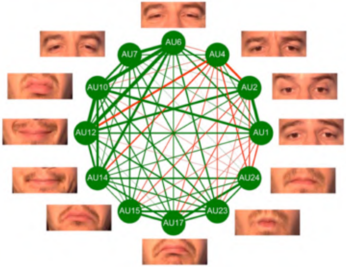
Table: Datasets used. † average number of labels per observation. ‡ number of labels per observation divided by the total number of labels, averaged over the samples.

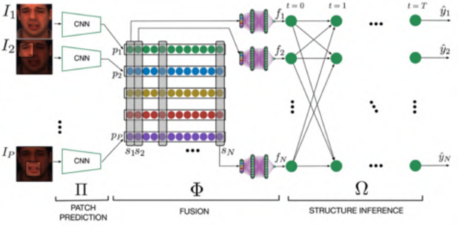
Ablation Study on BP4D



	method	AU01	AU02	AU04	AU06	AU07	AU10	AU12	AU14	AU15	AU17	AU23	AU24	avg
	Π(right eye)	38.0	[37.7]	48.3	69.5	71.0	72.4	77.4	50.7	15.0	38.9	13.8	15.3	45.7
	Π(between eye)	41.7	34.8	45.9	64.9	65.5	72.1	73.9	54.9	19.7	33.9	13.9	7.0	44.0
⊓	Π(mouth)	12.4	7.3	22.4	75.5	70.5	78.9	81.3	[66.2]	35.8	59.6	37.6	[42.8]	49.3
	Π(right cheek)	30.5	18.4	41.8	75.2	73.2	79.1	81.9	61.9	35.7	55.1	35.5	35.7	52.0
	Π(nose)	41.6	28.4	46.4	71.1	70.5	78.8	78.0	57.1	21.3	43.7	34.0	20.3	49.3
	Π(face)	43.8	37.5	[54.9]	**77.4**	71.2	79.2	[84.0]	56.6	39.7	59.7	[39.2]	39.5	56.9
	Π + Φ	[44.8]	35.8	[57.1]	[76.7]	**74.3**	[79.6]	83.7	56.6	**41.1**	[61.8]	**42.2**	40.1	[57.8]
	Π + Φ + Ω(DSIN)	**51.7**	**41.6**	**58.1**	76.6	[74.1]	**85.5**	**87.4**	**72.6**	[40.4]	**66.5**	38.6	**46.9**	**61.7**

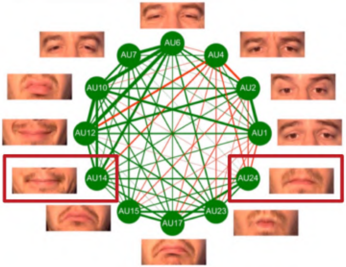
Ablation Study on BP4D: General

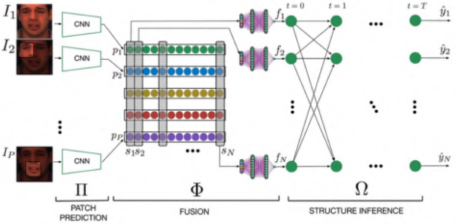




	method	AU01	AU02	AU04	AU06	AU07	AU10	AU12	AU14	AU15	AU17	AU23	AU24	avg
Π	Π(right eye)	38.0	[37.7]	48.3	69.5	71.0	72.4	77.4	50.7	15.0	38.9	13.8	15.3	45.7
	Π(between eye)	41.7	34.8	45.9	64.9	65.5	72.1	73.9	54.9	19.7	33.9	13.9	7.0	44.0
	Π(mouth)	12.4	7.3	22.4	75.5	70.5	78.9	81.3	[66.2]	35.8	59.6	37.6	[42.8]	49.3
	Π(right cheek)	30.5	18.4	41.8	75.2	73.2	79.1	81.9	61.9	35.7	55.1	35.5	35.7	52.0
	Π(nose)	41.6	28.4	46.4	71.1	70.5	78.8	78.0	57.1	21.3	43.7	34.0	20.3	49.3
	Π(face)	43.8	37.5	[54.9]	**77.4**	71.2	79.2	[84.0]	56.6	39.7	59.7	[39.2]	39.5	56.9
	Π + Φ	[44.8]	35.8	[57.1]	[76.7]	**74.3**	[79.6]	83.7	56.6	**41.1**	[61.8]	**42.2**	40.1	[57.8]
	Π + Φ + Ω(DSIN)	**51.7**	**41.6**	**58.1**	76.6	[74.1]	**85.5**	**87.4**	**72.6**	[40.4]	**66.5**	38.6	**46.9**	**61.7**

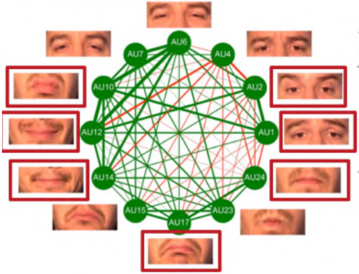
Ablation Study on BP4D: Patch Prediction

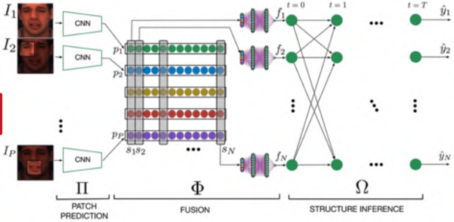




	method	AU01	AU02	AU04	AU06	AU07	AU10	AU12	AU14	AU15	AU17	AU23	AU24	avg
Π	Π(right eye)	38.0	[37.7]	48.3	69.5	71.0	72.4	77.4	50.7	15.0	38.9	13.8	15.3	45.7
	Π(between eye)	41.7	34.8	45.9	64.9	65.5	72.1	73.9	54.9	19.7	33.9	13.9	7.0	44.0
	Π(mouth)	12.4	7.3	22.4	75.5	70.5	78.9	81.3	[66.2]	35.8	59.6	37.6	[42.8]	49.3
	Π(right cheek)	30.5	18.4	41.8	75.2	73.2	79.1	81.9	61.9	35.7	55.1	35.5	35.7	52.0
	Π(nose)	41.6	28.4	46.4	71.1	70.5	78.8	78.0	57.1	21.3	43.7	34.0	20.3	49.3
	Π(face)	43.8	37.5	[54.9]	**77.4**	71.2	79.2	[84.0]	56.6	39.7	59.7	[39.2]	39.5	56.9
	$\Pi + \Phi$	[44.8]	35.8	[57.1]	[76.7]	**74.3**	[79.6]	83.7	56.6	**41.1**	[61.8]	**42.2**	40.1	[57.8]
	$\Pi + \Phi + \Omega$(DSIN)	**51.7**	**41.6**	**58.1**	76.6	[74.1]	**85.5**	**87.4**	**72.6**	[40.4]	**66.5**	38.6	**46.9**	**61.7**

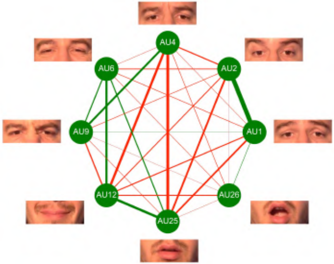
Ablation Study on BP4D: Structure Inference

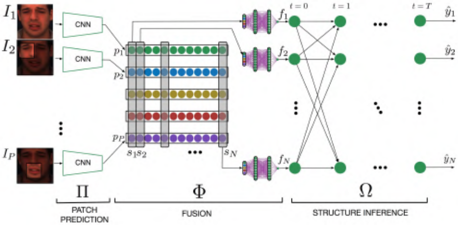




	method	AU01	AU02	AU04	AU06	AU07	AU10	AU12	AU14	AU15	AU17	AU23	AU24	avg
Π	Π(right eye)	38.0	[37.7]	48.3	69.5	71.0	72.4	77.4	50.7	15.0	38.9	13.8	15.3	45.7
	Π(between eye)	41.7	34.8	45.9	64.9	65.5	72.1	73.9	54.9	19.7	33.9	13.9	7.0	44.0
	Π(mouth)	12.4	7.3	22.4	75.5	70.5	78.9	81.3	[66.2]	35.8	59.6	37.6	[42.8]	49.3
	Π(right cheek)	30.5	18.4	41.8	75.2	73.2	79.1	81.9	61.9	35.7	55.1	35.5	35.7	52.0
	Π(nose)	41.6	28.4	46.4	71.1	70.5	78.8	78.0	57.1	21.3	43.7	34.0	20.3	49.3
	Π(face)	43.8	37.5	[54.9]	**77.4**	71.2	79.2	[84.0]	56.6	39.7	59.7	[39.2]	39.5	56.9
	Π + Φ	[44.8]	35.8	[57.1]	[76.7]	**74.3**	[79.6]	83.7	56.6	**41.1**	[61.8]	**42.2**	40.1	[57.8]
	Π + Φ + Ω(DSIN)	**51.7**	**41.6**	**58.1**	76.6	[74.1]	**85.5**	**87.4**	**72.6**	[40.4]	**66.5**	38.6	**46.9**	**61.7**

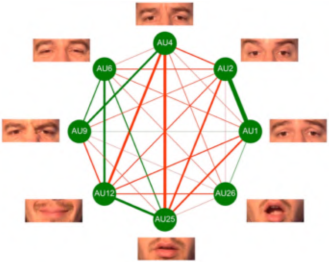
Ablation Study on DISFA

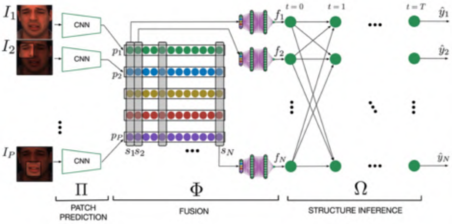




method	AU01	AU02	AU04	AU06	AU09	AU12	AU25	AU26	avg
Π(right eye)	27.2	15.4	58.8	8.0	18.2	53.6	73.3	9.1	33.0
Π(between eye)	34.6	13.2	59.7	15.4	21.1	50.9	72.9	8.5	34.5
Π(mouth)	7.5	6.4	44.6	28.5	23.9	**72.1**	87.5	[27.3]	37.2
Π(right cheek)	24.6	12.2	46.1	31.2	45.2	71.5	84.5	22.4	33.8
Π(nose)	21.9	19.1	52.0	**32.0**	**50.9**	66.5	76.6	8.9	41.0
Π(face)	29.8	[31.4]	64.6	26.8	21.3	70.1	87.0	20.3	43.9
$\Pi + \Phi$	[40.1]	18.6	**70.8**	25.4	42.1	[71.8]	[88.8]	26.4	[48.0]
$\Pi + \Phi + \Omega$(DSIN)	**42.4**	**39.0**	[68.4]	[28.6]	[46.8]	70.8	**90.4**	**42.2**	**53.6**

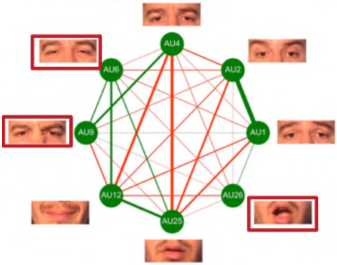
Ablation Study on DISFA: General

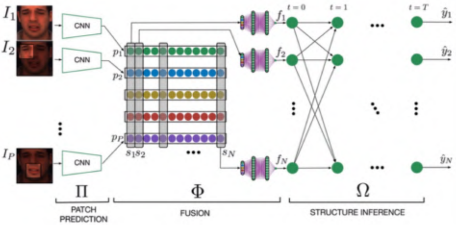




method	AU01	AU02	AU04	AU06	AU09	AU12	AU25	AU26	avg
Π(right eye)	27.2	15.4	58.8	8.0	18.2	53.6	73.3	9.1	33.0
Π(between eye)	34.6	13.2	59.7	15.4	21.1	50.9	72.9	8.5	34.5
Π(mouth)	7.5	6.4	44.6	28.5	23.9	**72.1**	87.5	[27.3]	37.2
Π(right cheek)	24.6	12.2	46.1	31.2	45.2	71.5	84.5	22.4	33.8
Π(nose)	21.9	19.1	52.0	**32.0**	**50.9**	66.5	76.6	8.9	41.0
Π(face)	29.8	[31.4]	64.6	26.8	21.3	70.1	87.0	20.3	43.9
$\Pi + \Phi$	[40.1]	18.6	**70.8**	25.4	42.1	[71.8]	[88.8]	26.4	[48.0]
$\Pi + \Phi + \Omega$(DSIN)	**42.4**	**39.0**	[68.4]	[28.6]	[46.8]	70.8	**90.4**	**42.2**	**53.6**

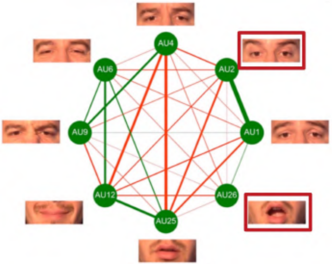
Ablation Study on DISFA: Patch Prediction

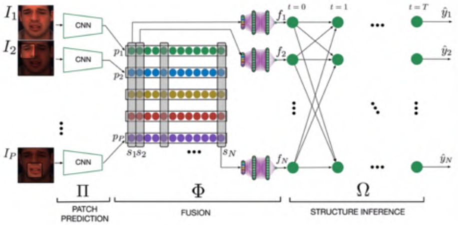




method	AU01	AU02	AU04	AU06	AU09	AU12	AU25	AU26	avg
Π(right eye)	27.2	15.4	58.8	8.0	18.2	53.6	73.3	9.1	33.0
Π(between eye)	34.6	13.2	59.7	15.4	21.1	50.9	72.9	8.5	34.5
Π(mouth)	7.5	6.4	44.6	28.5	23.9	**72.1**	87.5	[27.3]	37.2
Π(right cheek)	24.6	12.2	46.1	31.2	45.2	71.5	84.5	22.4	33.8
Π(nose)	21.9	19.1	52.0	**32.0**	**50.9**	66.5	76.6	8.9	41.0
Π(face)	29.8	[31.4]	64.6	26.8	21.3	70.1	87.0	20.3	43.9
$\Pi + \Phi$	[40.1]	18.6	**70.8**	25.4	42.1	[71.8]	[88.8]	26.4	[48.0]
$\Pi + \Phi + \Omega$(DSIN)	**42.4**	**39.0**	[68.4]	[28.6]	[46.8]	70.8	**90.4**	**42.2**	**53.6**

Ablation Study on DISFA: Structure Inference





method	AU01	AU02	AU04	AU06	AU09	AU12	AU25	AU26	avg
Π(right eye)	27.2	15.4	58.8	8.0	18.2	53.6	73.3	9.1	33.0
Π(between eye)	34.6	13.2	59.7	15.4	21.1	50.9	72.9	8.5	34.5
Π(mouth)	7.5	6.4	44.6	28.5	23.9	**72.1**	87.5	[27.3]	37.2
Π(right cheek)	24.6	12.2	46.1	31.2	45.2	71.5	84.5	22.4	33.8
Π(nose)	21.9	19.1	52.0	**32.0**	**50.9**	66.5	76.6	8.9	41.0
Π(face)	29.8	[31.4]	64.6	26.8	21.3	70.1	87.0	20.3	43.9
$\Pi + \Phi$	[40.1]	18.6	**70.8**	25.4	42.1	[71.8]	[88.8]	26.4	[48.0]
$\Pi + \Phi + \Omega$(DSIN)	**42.4**	**39.0**	[68.4]	[28.6]	[46.8]	70.8	**90.4**	**42.2**	**53.6**

Comparison with State-of-the-Art on BP4D

method	AU01	AU02	AU04	AU06	AU07	AU10	AU12	AU14	AU15	AU17	AU23	AU24	AVG
JPML[10]	32.6	25.6	37.4	42.3	50.5	72.2	74.1	[65.7]	38.1	40.0	30.4	[42.3]	45.9
DRML[11]	36.4	**41.8**	43.0	55.0	67.0	66.3	65.8	54.1	33.2	48.0	31.7	30.0	48.3
CPM[12]	[43.4]	40.7	43.3	59.2	61.3	62.1	68.5	52.5	36.7	54.3	**39.5**	37.8	50.0
ROI[13]	36.2	31.6	43.4	**77.1**	[73.7]	[85.0]	[87.0]	62.6	**45.7**	58.0	38.3	37.4	56.4
DSIN	**51.7**	[41.6]	**58.1**	[76.6]	**74.1**	**85.5**	**87.4**	**72.6**	[40.4]	**66.5**	38.6	**46.9**	**61.7**

Table: AU recognition results on BP4D. Best results are shown in bold. Second best results are shown in brackets.

[10] K. Zhao, W.-S. Chu, F. De la Torre, J. F. Cohn, and H. Zhang. Joint patch and multi-label learning for facial action unit detection. In Proceedings of the IEEE CVPR, pages 2207–2216, 2015.

[11] K. Zhao, W.-S. Chu, and H. Zhang. Deep region and multi-label learning for facial action unit detection. In Proceedings of the IEEE CVPR, pages 3391–3399, 2016.

[12] J. Zeng, W.-S. Chu, F. De la Torre, J. F. Cohn, and Z. Xiong. Confidence preserving machine for facial action unit detection. In Proceedings of the IEEE ICCV pages 3622-3630, 2015.

[13] Li, Wei, Farnaz Abtahi, and Zhigang Zhu. "Action unit detection with region adaptation, multi-labeling learning and optimal temporal fusing." Proceedings of the IEEE CVPR. 2017.

Comparison with State-of-the-art on DISFA

method	AU01	AU02	AU04	AU06	AU09	AU12	AU25	AU26	avg
APL[14]	11.4	12.0	30.1	12.4	10.1	65.9	21.4	[26.0]	23.8
DRML[15]	17.3	17.7	37.4	29.0	[10.7]	37.7	38.5	20.1	26.7
ROI[16]	[41.5]	[26.4]	[66.4]	**50.7**	8.5	**89.3**	[88.9]	15.6	[48.5]
DSIN	**46.9**	**42.5**	**68.8**	[32.0]	**51.8**	[73.1]	**91.9**	**46.6**	**56.7**

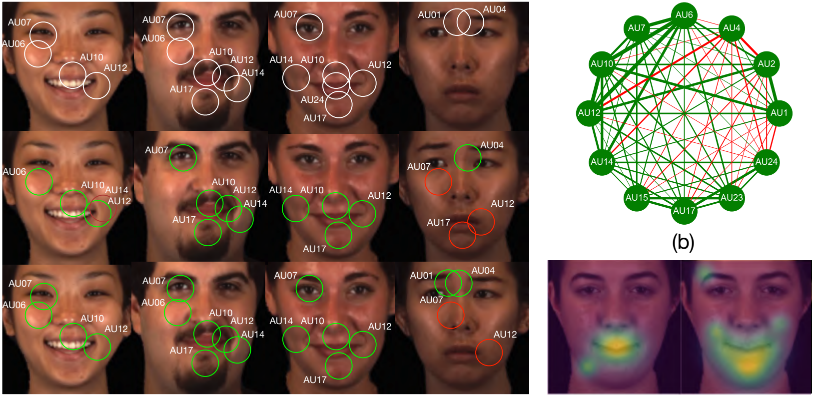
Table: AU recognition results on DISFA. Best results are shown in bold. Second best results are shown in brackets.

[14] L. Zhong, Q. Liu, P. Yang, J. Huang, and D. N. Metaxas. Learning multiscale active facial patches for expression analysis. IEEE Transactions on cybernetics, 45(8):1499–1510, 2015.

[15] K. Zhao, W.-S. Chu, and H. Zhang. Deep region and multi-label learning for facial action unit detection. In Proceedings of the IEEE CVPR, pages 3391–3399, 2016.

[16] Li, Wei, Farnaz Abtahi, and Zhigang Zhu. "Action unit detection with region adaptation, multi-labeling learning and optimal temporal fusing." Proceedings of the IEEE CVPR. 2017.

Qualitative Results



(a) (c)

Figure: (a) Examples of AU predictions: ground-truth (top), fusion module (middle) and structure inference (bottom) prediction (●: true positive, ●: false positive). (b) AUs correlation in BP4D (●: positive, ●: negative). (c) Class activation map for AU24 that shows the discriminative regions of simple patch prediction (left) and DSIN (right).

Outline I

DSIN Achieves High-performance

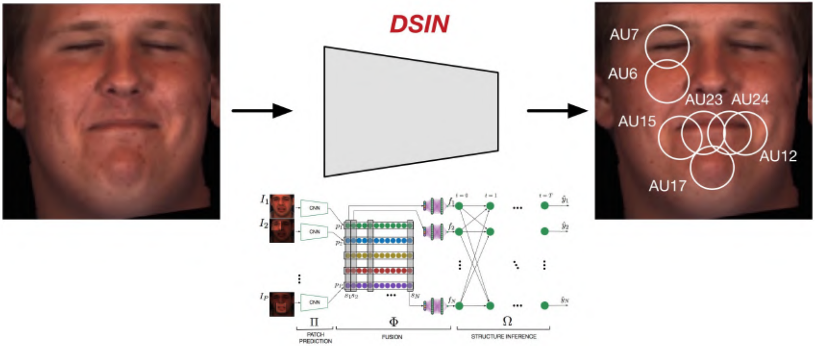


Figure: We have proposed DSIN, a DNN capable of recognizing AUs with state-of-the-art performance.

DSIN Is a Black-box Model

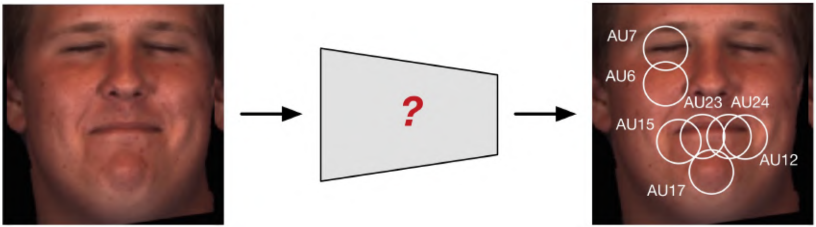


Figure: Paradoxically, even though we are DSIN's designers, its complexity makes it uninterpretable.

Approach

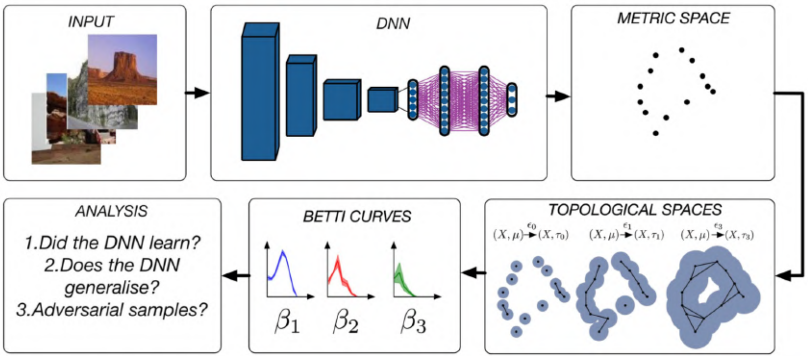


Figure: An overview of the approach used.

From DNNs to Metric Spaces

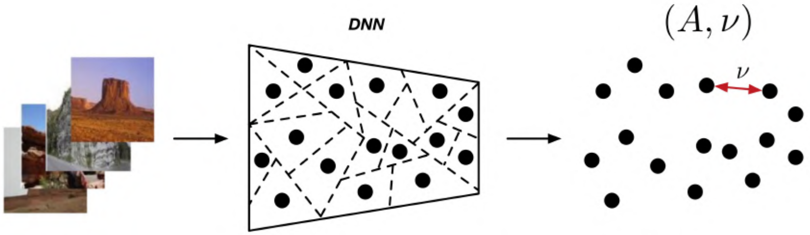


Figure: Projecting a DNN into a metric space.

Studying Metric Spaces with Algebraic Topology

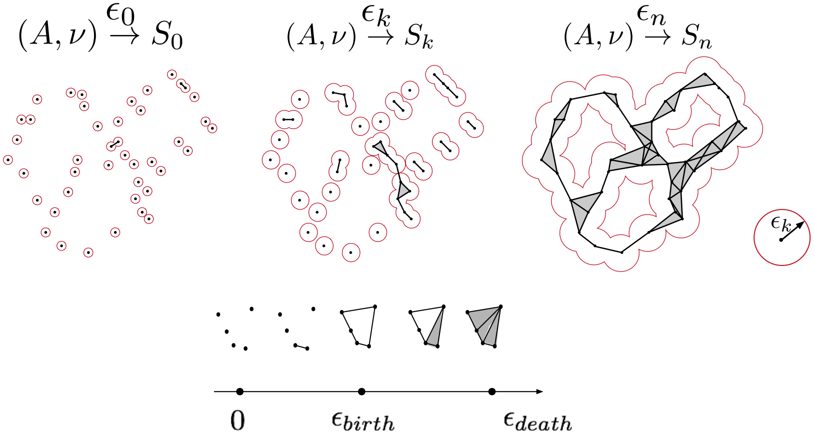
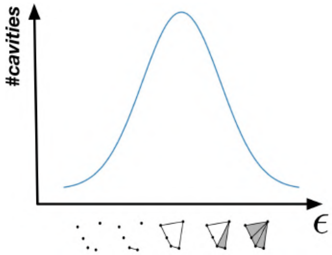


Figure: Given a metric space, the Vietoris-Rips filtration creates a nested sequence of simplicial complexes by connecting points situated closer than a predefined distance ϵ.

The Betti Curve



$$\beta_d = \{\sum_{S_k} \mathbf{1}_{cav}; k = 0, \ldots, n\}$$

Figure: The Betti Curve is a compact descriptor of topological objects.

From DNNs to Betti Curves

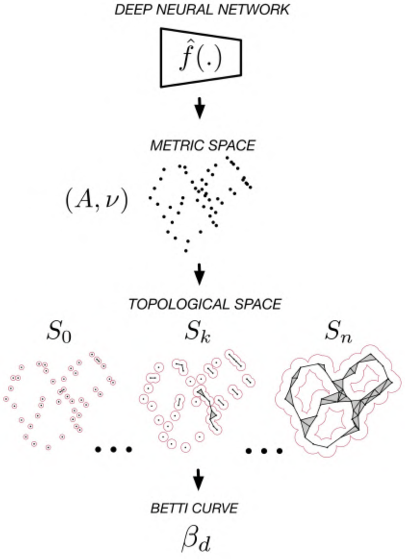


Figure: An overview of computing Betti curves from DNNs.

Learning vs. Memorization

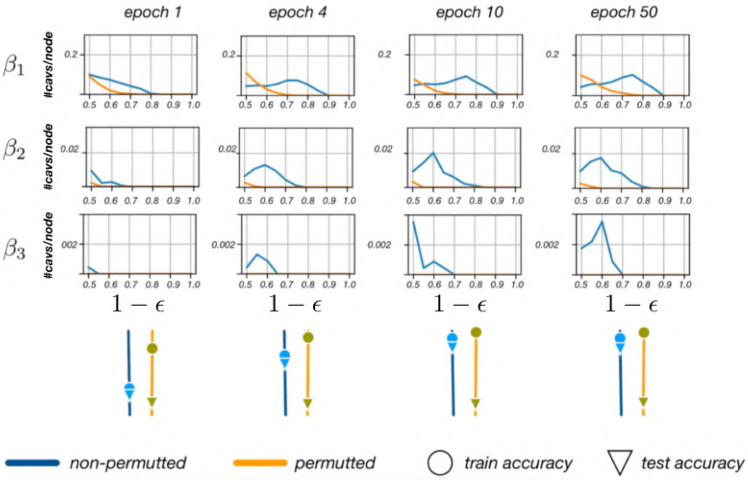


Figure: Betti curves for learning (blue) and memorization (orange). LeNet5 on MNIST.

Unaltered vs. Adversarial Attacks

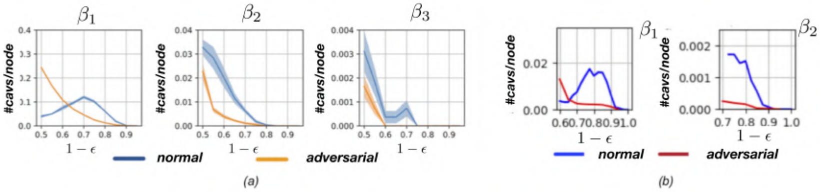


Figure: Betti curves obtained when using unaltered and adversarial testing samples for (a) LeNet5 on MNIST and (b) VGG16 on Imagenet[17].

[17] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard. "Deep-fool: a simple and accurate method to fool deep neural networks". In Proceedings of the IEEE CVPR, pages 2574–2582, 2016.

Learning and Generalization

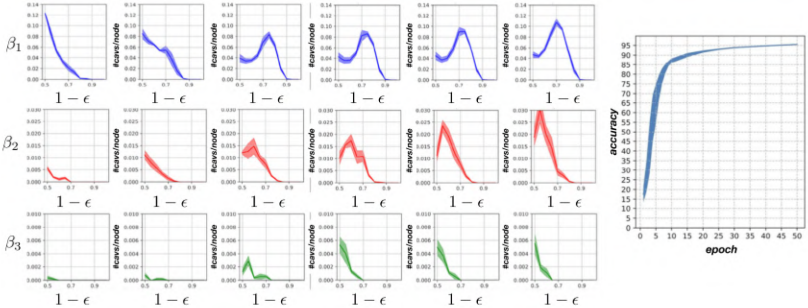


Figure: Betti number dynamics during LeNet5 training on MNIST (left); accuracy dynamic (right).

Main Theoretical Results[18]

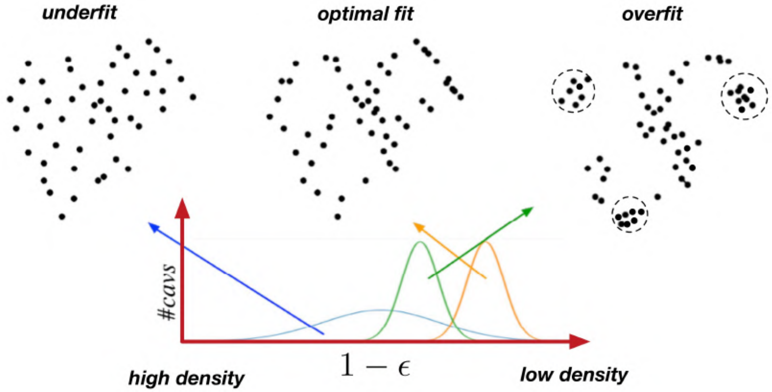


Figure: An illustration of corresponding metric spaces of underfitting (left), optimal fitting (centre) and overfitting (right).

[18] **Learning to generalize** in DNN is defined by the creation of 2D and 3D cavities in the topological space representing the correlations of activation of distant nodes of the DNN, and the movement of 1D cavities from higher to lower density. **Memorizing** (overfitting) is indicated by a regression of these cavities toward higher densities in the topological space.

Topological Early Stopping (TES)

Algorithm 1 Topological Early Stopping.

Input: train dataset $\mathcal{X} = \{x_i, y_i\}_{i=1}^{n}$; DNN partition A.

repeat

  $\omega \leftarrow arg\,min_{\omega} L(\hat{f}(x;\omega), y)$. ▷ Train DNN and estimate f by optimizing loss L over $\mathcal{X}$.

  for all pairs of nodes $(a_p, a_q) \in A, p \neq q$ **do**

    $\nu_{pq} \leftarrow \sum_{i=1}^{n} \frac{(a_{pi} - \overline{a_p})(a_{qi} - \overline{a_q})}{s_{a_p} s_{a_q}}$ ▷ Compute correlations.

  end for

  $\mathcal{S} \leftarrow VR(A, \nu)$ ▷ Perform Vietoris-Rips filtration and get set of simplicial complexes $\mathcal{S}$.

  $P \leftarrow \mathcal{PH}(\mathcal{S})$ ▷ Compute persistent homology $\mathcal{PH}$ over $\mathcal{S}$ and get persistent diagram P.

  $\beta_d \leftarrow \{\sum_{S_k} \mathbf{1}_{cav}, k = 1, \ldots, n\}$ ▷ Compute Betti Curves.

  $\widehat{k}_t = \arg\max_k \beta_d(S_k)$.

  $t \leftarrow t + 1$

until $\widehat{k}_t > \widehat{k}_{t-1}$.

Figure: Topological early stopping algorithm.

TES in Practice

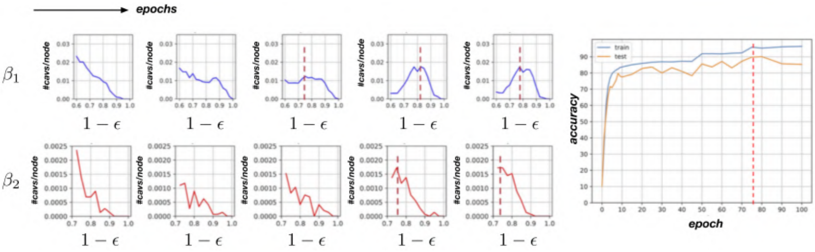


Figure: (Left) Betti numbers dynamic during VGG16 training on Imagenet; (Right) Accuracy dynamic.

TES on DSIN

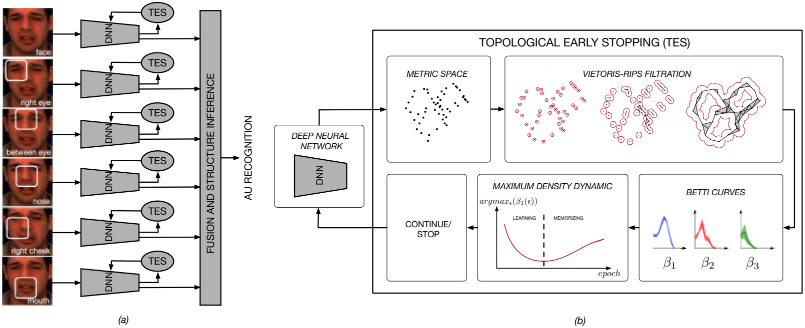


Figure: An overview of applying Topological Early Stopping on DSIN for AU recognition.

Experimental Results on BP4D

method	AU01	AU02	AU04	AU06	AU07	AU10	AU12	AU14	AU15	AU17	AU23	AU24	AVG
JPML	32.6	25.6	37.4	42.3	50.5	72.2	74.1	**65.7**	38.1	40.0	30.4	42.3	45.9
DRML	36.4	41.8	43.0	55.0	67.0	66.3	65.8	54.1	33.2	48.0	31.7	30.0	48.3
CPM	43.4	40.7	43.3	59.2	61.3	62.1	68.5	52.5	36.7	54.3	39.5	37.8	50.0
ROI	36.2	31.6	43.4	**77.1**	73.7	**85.0**	**87.0**	62.6	**45.7**	58.0	38.3	37.4	56.4
$DSIN_{40}^{ESP}$	[49.9]	41.2	54.1	73.1	73.4	80.0	[84.9]	[63.5]	35.2	63.1	42.1	41.6	58.5
$DSIN_{10}^{ESP}$	49.7	[42.5]	[56.6]	72.0	[74.7]	[81.1]	82.2	62.2	36.5	[63.9]	40.1	**43.3**	58.7
$DSIN^{ESNP}$	49.1	42.0	**57.8**	71.6	72.6	79.6	82.4	62.2	[43.1]	63.4	**46.1**	42.4	[59.4]
$DSIN^{TES}$	**50.4**	**44.3**	56.2	[73.3]	**75.6**	79.3	83.2	61.2	42.7	**65.2**	[44.2]	[43.1]	**59.9**

Figure: AU recognition results on BP4D. Best results are shown in bold. Second best results are shown in brackets.

Experimental Results on DISFA

method	AU01	AU02	AU04	AU06	AU09	AU12	AU25	AU26	avg
APL	11.4	12.0	30.1	12.4	10.1	65.9	21.4	26.0	23.8
DRML	17.3	17.7	37.4	29.0	10.7	37.7	38.5	20.1	26.7
ROI	41.5	26.4	[66.4]	**50.7**	8.5	**89.3**	88.9	15.6	48.5
$DSIN_{40}^{ESP}$	**45,3**	38.0	65.2	29.4	[42.8]	[73.8]	**90.2**	**41.5**	[53.3]
$DSIN_{10}^{ESP}$	43,2	39.1	**67.3**	31.2	42.6	73.5	[89.1]	40.3	[53.3]
$DSIN^{ESNP}$	41.4	**45.3**	61.4	[34.9]	39.1	70.5	87.0	37.9	52.2
$DSIN^{TES}$	[44.4]	[43.6]	64.8	33.1	**43.1**	72.2	88.0	[41.3]	**53.8**

Figure: AU recognition results on DISFA. Best results are shown in bold. Second best results are shown in brackets.

TES vs ESP[20]

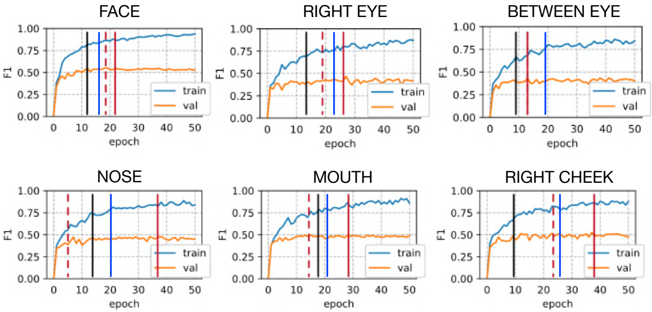


Figure: Stopping decision comparison per subnetwork of DSIN on the BP4D. TES decision (black line), ESP_{40} (red line), ESP_{10} (red dashed line) and ESNP[19] (blue line).

[19] Rieck, Bastian, et al. "Neural persistence: A complexity measure for deep neural networks using algebraic topology." arXiv preprint arXiv:1812.09764 (2018).

[20] Early Stopping with Patience

Qualitative Results

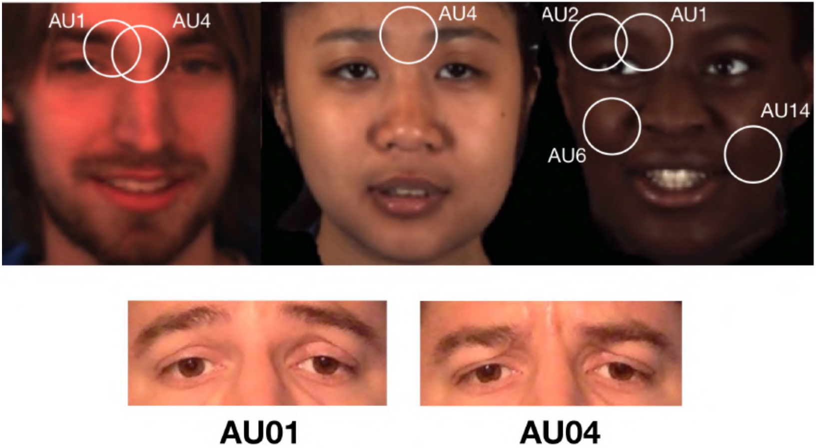


Figure: After a sufficient number of iterations, if the training continues, the generalization gap increases without any significant improvement on the validation set. We show here some examples of noisy labels that the dedicated networks of DSIN memorize between the epoch TES would stop and the epoch ESP would stop.

Outline I

Contributions

1. **Performance in Facial Expression Recognition.**
   1.1 Proposal of a model that learns representation, patch and output structure of the face end-to-end.
   1.2 Introduction of a structure inference topology that replicates inference algorithm in probabilistic graphical models by using a recurrent neural network.
   1.3 Extended ablation study and experimental analysis of the newly proposed architecture.
2. **Interpretability in Facial Expression Recognition.**
   2.1 Formulation of novel general framework for analysis of deep neural networks based on algebraic topology.
   2.2 Analysis of fundamental topological differences between DNNs that learn and DNNs that memorize.
   2.3 Analyze and improving performance of the previously proposed architecture for facial expression architecture using the new theoretical framework.

Topology Correlates with Performance Gap [21]

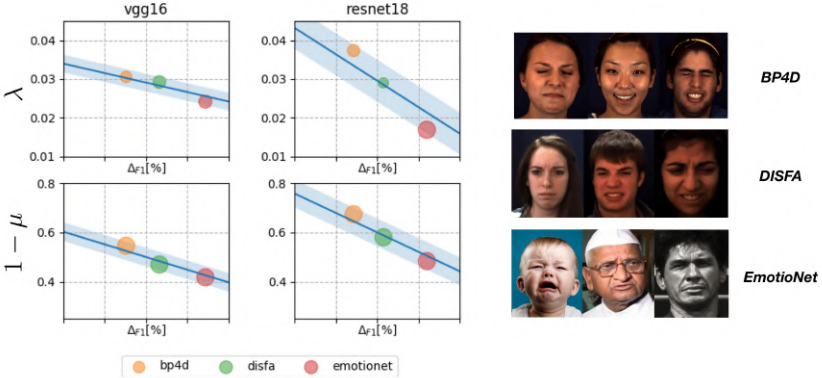


Figure: Topology summaries against performance (accuracy) gap for different models trained to recognize objects. Each disc represents mean (centre) and standard deviation (radius) on a particular dataset. Linear mapping and the corresponding standard deviation of the observed samples are marked.

[21] Under revision

Experimental Results for Action Unit Recognition

Model	resnet18	vgg16	mean
$g(\lambda, \mu)$	5.18±3.62	5.87±3.62	5.52±3.62

Table: Evaluation for AU recognition. Mean and standard deviation error (in %) in estimating the test performance.

Journals

- Corneanu, C. A., Simón, M. O., Cohn, J. F., Guerrero, S. E. (2016). Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. IEEE TPAMI, 38(8), 1548-1568.
- Kulkarni, Kaustubh, Ciprian Corneanu, Ikechukwu Ofodile, Sergio Escalera, Xavier Baro, Sylwia Hyniewska, Juri Allik, and Gholamreza Anbarjafari. "Automatic recognition of facial displays of unfelt emotions." IEEE TAC (2018).
- Noroozi, Fatemeh, Ciprian Corneanu, Dorota Kaminska, Tomasz Sapinski, Sergio Escalera, and Gholamreza Anbarjafari. "Survey on emotional body gesture recognition." IEEE TAC (2018).
- Simón, Marc Oliu, Ciprian Corneanu, Kamal Nasrollahi, Olegs Nikisins, Sergio Escalera, Yunlian Sun, Haiqing Li, Zhenan Sun, Thomas B. Moeslund, and Modris Greitans. "Improved RGB-DT based face recognition." Iet Biometrics 5, no. 4 (2016): 297-303.

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

IEEE TRANSACTIONS ON AFFECTIVE COMPUTING

International Conferences and Workshops

- Corneanu, Ciprian, Meysam Madadi, and Sergio Escalera. "Deep structure inference network for facial action unit recognition." In Proceedings of the ECCV, pp. 298-313. 2018.
- Corneanu, Ciprian A., Meysam Madadi, Sergio Escalera, and Aleix M. Martinez. "What Does It Mean to Learn in Deep Networks? And, How Does One Detect Adversarial Attacks?." In Proceedings of the IEEE CVPR, pp. 4757-4766. 2019.
- Oliu, Marc, Ciprian Corneanu, László A. Jeni, Jeffrey F. Cohn, Takeo Kanade, and Sergio Escalera. "Continuous supervised descent method for facial landmark localisation." In Proceedings of the Asian Conference on Computer Vision (ACCV), pp. 121-135. Springer, Cham, 2016.

International Conferences and Workshops

- Irani, Ramin, Kamal Nasrollahi, Marc O. Simon, Ciprian A. Corneanu, Sergio Escalera, Chris Bahnsen, Dennis H. Lundtoft et al. "Spatiotemporal analysis of RGB-DT facial images for multimodal pain level recognition." In Proceedings of the IEEE CVPRW, pp. 88-95. 2015.
- Escalera, Sergio, Mercedes Torres Torres, Brais Martinez, Xavier Baró, Hugo Jair Escalante, Isabelle Guyon, Georgios Tzimiropoulos et al. "Chalearn looking at people and faces of the world: Face analysis workshop and challenge 2016." In Proceedings of the IEEE CVPRW, pp. 1-8. 2016.
- Ponce-López, Víctor, Baiyu Chen, Marc Oliu, Ciprian Corneanu, Albert Clapés, Isabelle Guyon, Xavier Baró, Hugo Jair Escalante, and Sergio Escalera. "Chalearn lap 2016: First round challenge on first impressions-dataset and results." In Proceedings of the ECCV, pp. 400-418. Springer, Cham, 2016.

"The face is the mirror of the mind, and eyes without speaking confess the secrets of the heart."
St. Jerome

Thank you!