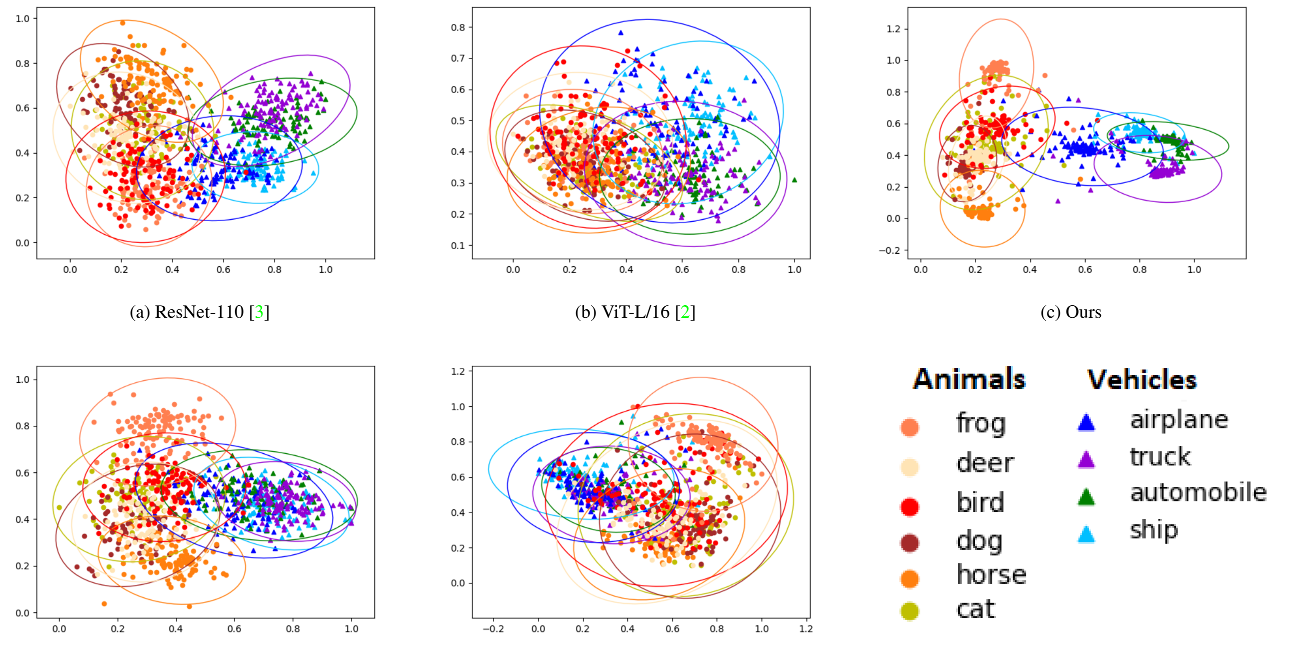
Supplementary material: Interpretable part-whole hierarchies and conceptual-semantic relationships in neural networks



(a) ResNet-110 [3]
(b) ViT-L/16 [2]
(c) Ours
(d) ConvMLP-L [5]
(e) Matrix-CapsNet [4]
(f) Legend

Figure 1. 2D representation of the latent space for multiple methods trained only on the CIFAR-10 dataset obtained using Principal Component Analysis (PCA) [7]. The PCA provides a deterministic change of base for the data from a multidimensional space into a 2D space. The legend (f) displays the classes, which are divided between super-classes *Vehicles* and *Animals* following the WordNet hierarchy [6]. The different methods (a,b,c,d,e) are all able to cluster the samples. However, while (a,b,e) display a latent space where classes are close to each other, the two MLP-based methods (c,d) are able to provide a clearer separation between classes. Both methods show conceptual-semantically close samples on the edge of each superclass, such as airplanes and birds. Inside each superclass, semantically close samples are represented contiguously, such as deers and horses, or cars and trucks. Our method (c) provides better inter-class and intra-class separability. We provide numerical results of the classes overlap in Fig. 2.

A. Content

In the supplementary material we attached:

- Further qualitative results about the latent space organization as the representation of conceptual-semantic relationship in data shown in Fig. 1.
- Additional results in Fig. 2 comparing the correlation between two classes in the latent space.
- A .zip file with the code to run and reproduce our experimental setup and results. The code is written using PyTorch Lightning and it will be included in a GitHub repository upon acceptance.

ResNet-110	Frog	Deer	Bird	Dog	Horse	Cat	Airplane	Truck	Automobile	Ship
Frog	100	32	70	12	3	30	24	0	0	3
Deer	32	100	42	50	22	73	14	1	1	0
Bird	70	42	100	18	7	39	27	3	3	6
Dog	12	50	18	100	39	52	5	0	0	0
Horse	3	22	7	39	100	33	4	4	4	0
Cat	30	73	39	52	33	100	17	3	4	2
Airplane	24	14	27	5	4	17	100	19	21	44
Truck	0	1	3	0	4	3	19	100	70	15
Automobile	0	1	3	0	4	4	21	70	100	18
Ship	3	0	6	0	0	2	44	15	18	100

(a) ResNet-110 [3]

ViT-L/16	Frog	Deer	Bird	Dog	Horse	Cat	Airplane	Truck	Automobile	Ship
Frog	100	75	55	73	59	73	25	18	13	14
Deer	75	100	71	56	53	58	30	18	13	18
Bird	55	71	100	48	56	52	39	22	16	22
Dog	73	56	48	100	65	82	20	19	16	11
Horse	59	53	56	65	100	65	27	28	24	20
Cat	73	58	52	82	65	100	24	23	20	14
Airplane	25	30	39	20	27	24	100	42	35	60
Truck	18	18	22	19	28	23	42	100	71	41
Automobile	13	13	16	16	24	20	35	71	100	35
Ship	14	18	22	11	20	14	60	41	35	100

(b) ViT-L/16 [2]

Ours	Frog	Deer	Bird	Dog	Horse	Cat	Airplane	Truck	Automobile	Ship
Frog	100	2	17	0	0	17	0	0	0	0
Deer	2	100	26	63	14	29	0	0	0	0
Bird	17	26	100	19	0	54	12	0	0	0
Dog	0	63	19	100	14	27	0	0	0	0
Horse	0	14	0	14	100	14	0	0	0	0
Cat	17	29	54	27	14	100	10	0	0	0
Airplane	0	0	12	0	0	10	100	18	11	16
Truck	0	0	0	0	0	0	18	100	19	11
Automobile	0	0	0	0	0	0	11	19	100	44
Ship	0	0	0	0	0	0	16	11	44	100

(c) Ours

ConvMLP-L	Frog	Deer	Bird	Dog	Horse	Cat	Airplane	Truck	Automobile	Ship
Frog	100	6	23	5	0	18	9	0	3	0
Deer	6	100	42	85	33	58	18	0	9	3
Bird	23	42	100	39	8	62	35	5	21	11
Dog	5	85	39	100	30	57	17	0	9	3
Horse	0	33	8	30	100	16	6	0	1	0
Cat	18	58	62	57	16	100	23	1	13	5
Airplane	9	18	35	17	6	23	100	38	61	53
Truck	0	0	5	0	0	1	38	100	59	70
Automobile	3	9	21	9	1	13	61	59	100	66
Ship	0	3	11	3	0	5	53	70	66	100

(d) ConvMLP-L [5]

Matrix-CapsNet	Frog	Deer	Bird	Dog	Horse	Cat	Airplane	Truck	Automobile	Ship
Frog	100	21	5	15	25	50	19	60	73	18
Deer	21	100	28	66	74	16	77	24	28	57
Bird	5	28	100	25	33	6	31	6	8	10
Dog	15	66	25	100	58	10	75	16	21	54
Horse	25	74	33	58	100	22	76	31	32	44
Cat	50	16	6	10	22	100	15	69	49	9
Airplane	19	77	31	75	76	15	100	21	25	48
Truck	60	24	6	16	31	69	21	100	70	18
Automobile	73	28	8	21	32	49	25	70	100	25
Ship	18	57	10	54	44	9	48	18	25	100

(e) Matrix-CapsNet [4]

Figure 2. The overlap percentage O between classes in the latent space is reported for each possible class and each method. For each table, the top-left quadrant represents the overlap percentage between classes belonging to the super-class *animals*, while the bottom-right one for the superclass *vehicles*. The top-right and bottom-left quadrants represent the area where a mistake with a higher hierarchical severity [1] is possible. It would be ideal to have the table with zeros for all the values, but the diagonal. That would represent perfect separation between all the classes. Our method provides the best separation between the two superclasses. It is interesting to notice that the highest intra-superclasses correlation is present between the two classes *bird* and *airplanes* which share features like the wings and the ability to fly.

References

[1] Luca Bertinetto, Romain Mueller, Konstantinos Tertikas, Sina Samangooei, and Nicholas A Lord. Making better mistakes: Leveraging class hierarchies with deep networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12506–12515, 2020. 2

[2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 1, 2

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 2

[4] Geoffrey E Hinton, Sara Sabour, and Nicholas Frosst. Matrix capsules with em routing. In *International conference on learning representations*, 2018. 1, 2

[5] Jiachen Li, Ali Hassani, Steven Walton, and Humphrey Shi. Convmlp: Hierarchical convolutional mlps for vision. *arXiv preprint arXiv:2109.04454*, 2021. 1, 2

[6] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995. 1

[7] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987. 1