Improving Latent Representations of ConvNets for Visual Understanding

Amélioration des représentations latentes des ConvNets pour l'interprétation de données visuelles

Thomas Robert – 3 octobre 2019

Jury de These

Rapporteurs	Examinateurs	Encadrants
Stéphane Canu	Catherine Achard	Matthieu Cord
Greg Mori	Kartheek Alahari	Nicolas Thome
	David Picard	

Machine Learning & Deep Learning for Information Access

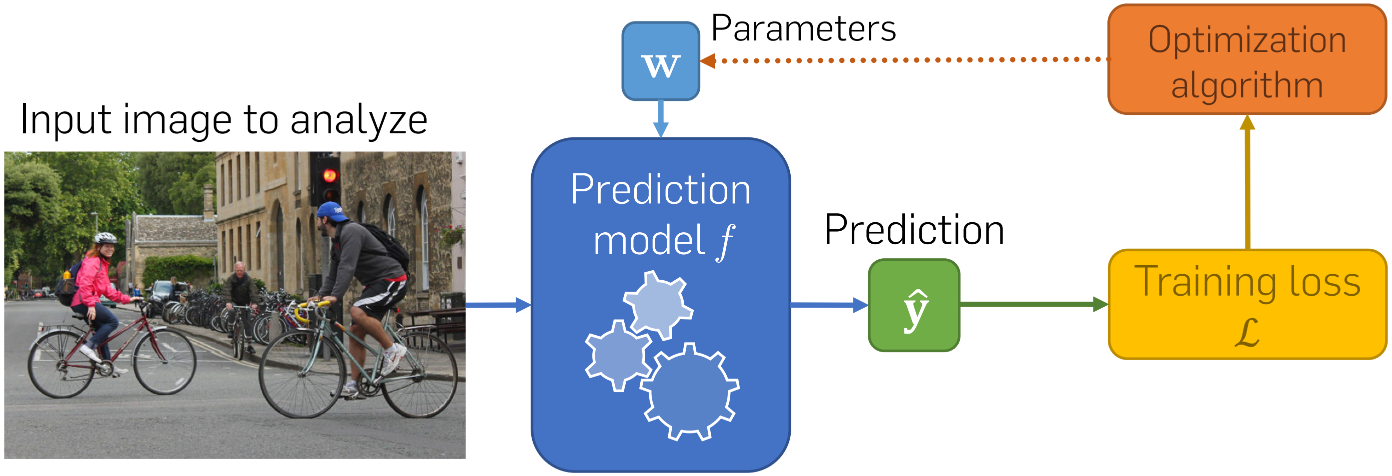
Visual content in the digital era

SCIENCES SORBONNE UNIVERSITÉ

- Exponential increase in quantity of images/videos taken across the world
  - Youtube: 500h of video / min
  - Facebook: 300M photos / day

- How to extract semantic information?

Computer Vision and Machine Learning

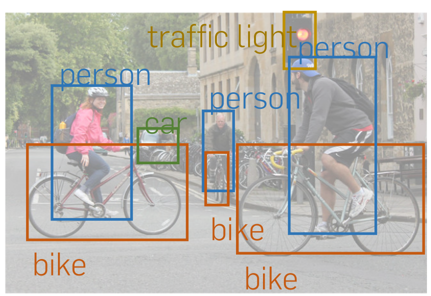


Tasks

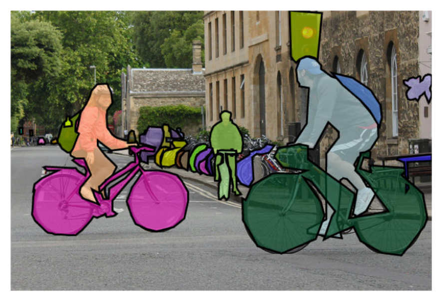
Classification

- ✓ Street
- ✗ Office
- ✗ Bedroom
- ✗ …

Object detection



Segmentation



Captioning

a girl in a pink jacket on a bicycle passes a man in a blue cap on a bicycle.

…

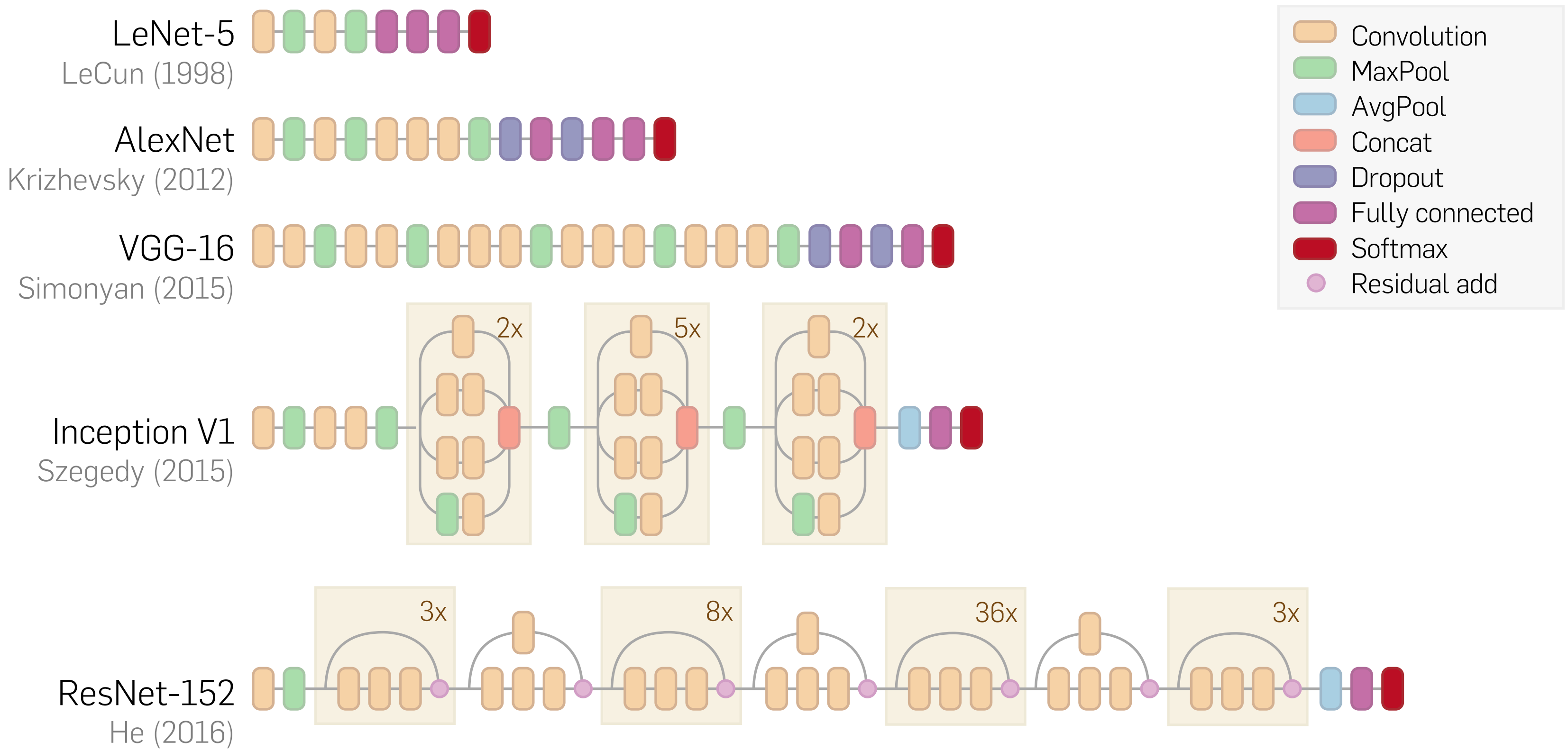
Defect detection

Photo search

Autonomous driving

Visually impaired assistant

Deep Learning



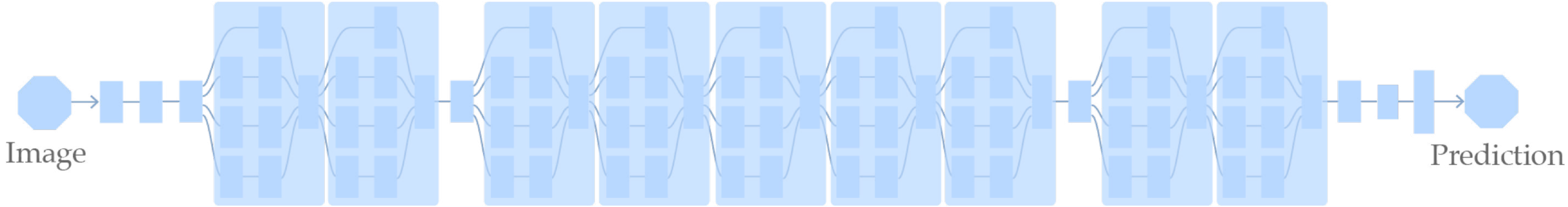
more depth → more parameters → more data

Convolutions Skip-connections	10-100M	ImageNet 1.3M images 1000 classes

Typical deep learning model

Deep ConvNet = succession of latent representations

No semantic structure a priori

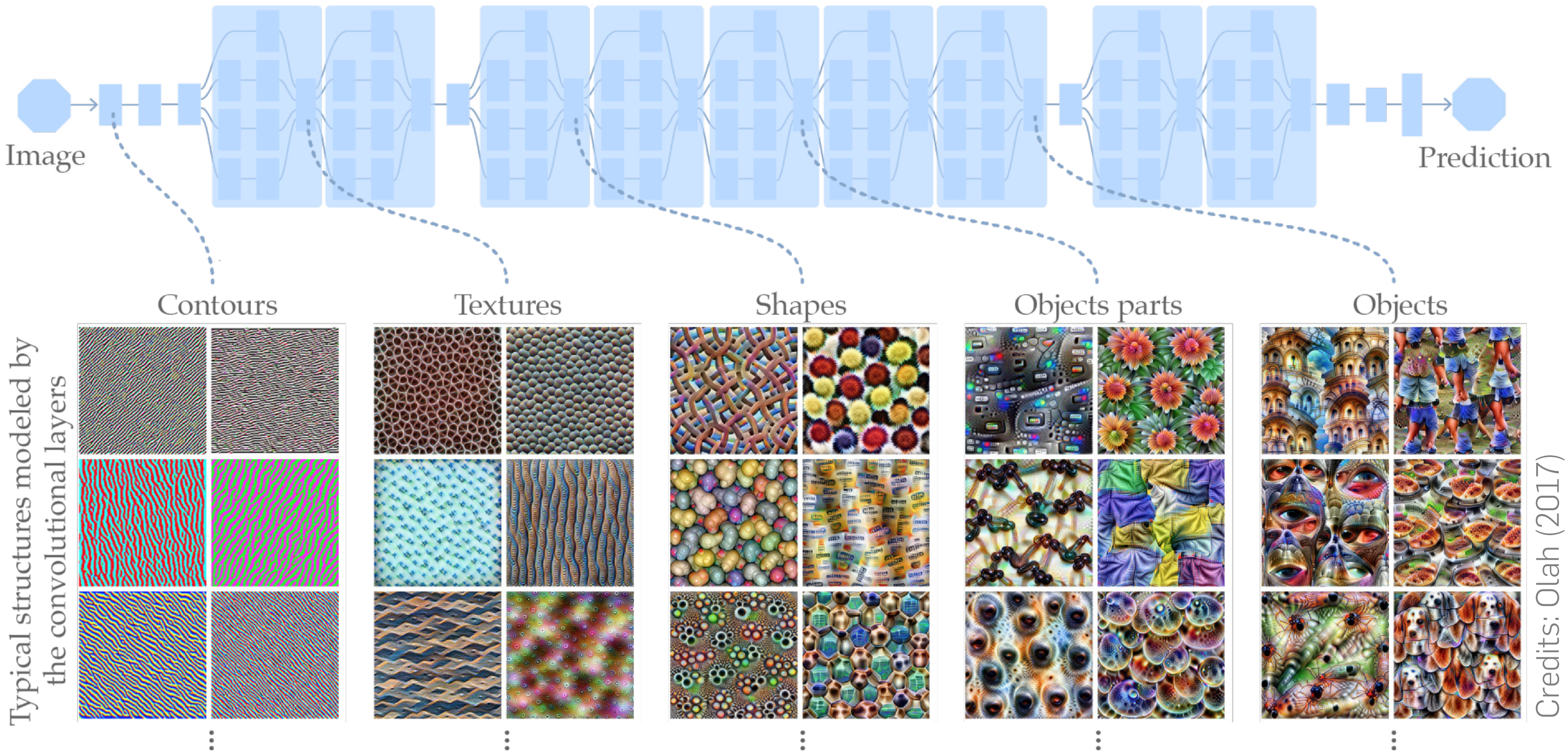


Typical deep learning model

SCIENCES SORBONNE UNIVERSITÉ

Deep ConvNet = succession of latent representations

No semantic structure a priori

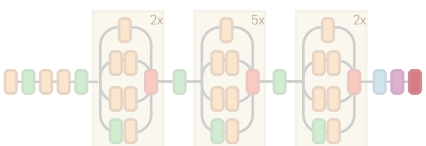


Semantic structure a posteriori (not really usable)

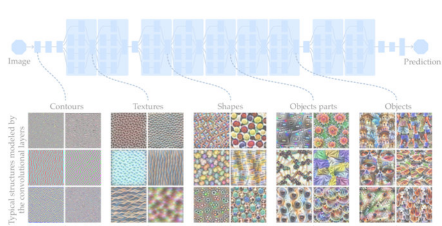
Motivations and contributions

Motivations

Models require **large labeled datasets**



Latent representations are **unstructured**



Axes

Regularization
generalize better with less data

Semi-supervised learning
use unlabeled data

Information structuration
in two branches

Disentangling
of semantic factors

Contribution

1 SHADE

2 HybridNet

3 DualDis

Training and improving deep ConvNets

$$\min_{\mathbf{w}} \mathcal{L}(\mathcal{D}, \mathbf{w}) = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y}) \in \mathcal{D}} \left[\mathcal{L}_{\text{task}}(f_{\mathbf{w}}(\mathbf{x}), \mathbf{y}) + \Omega_{\text{regul}}(\mathbf{w}, \mathbf{x}, \mathbf{y}) \right]$$

How to improve the generalization performance?

Data $\mathcal{D}$

- Data augmentation (3)
- Noise injection
- Semi-supervised learning (2)
- ...

Model f

- Convolutions
- Invariance
- Dropout
- Batch Norm
- ...

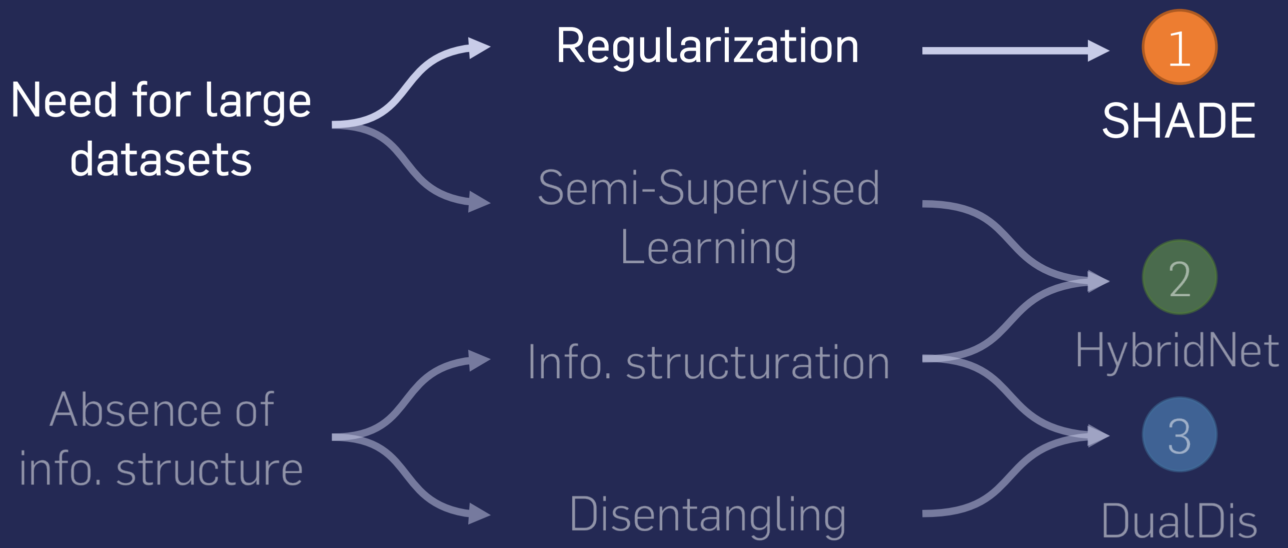
Regul. loss Ω_{regul}

- Weight decay
- Stability (2)
- Reconstruction (2)
- Entropy (1)
- ...

Regularization

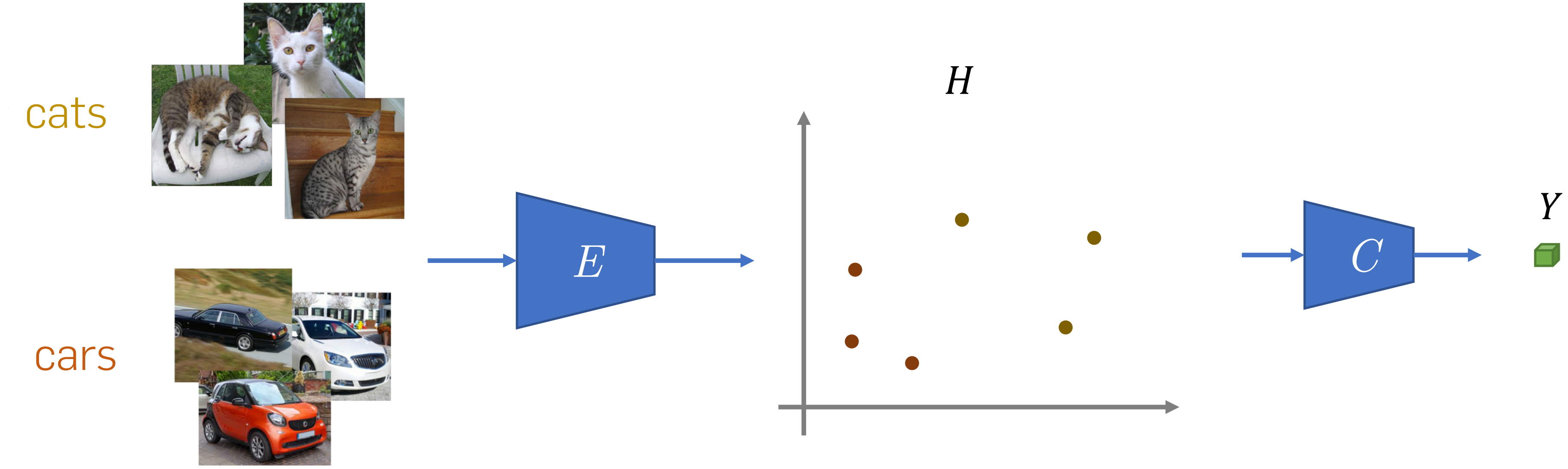
SHADE: Invariance through Conditional Entropy Minimization

SHADE: Information-Based Regularization for Deep Learning
Michael Blot, Thomas Robert, Nicolas Thome and Matthieu Cord
ICIP 2018, Best paper award



Classification, invariance and entropy

$$\min_{\mathbf{w}} \mathcal{L}_{\text{classif}}(\hat{\mathbf{y}}, \mathbf{y})$$



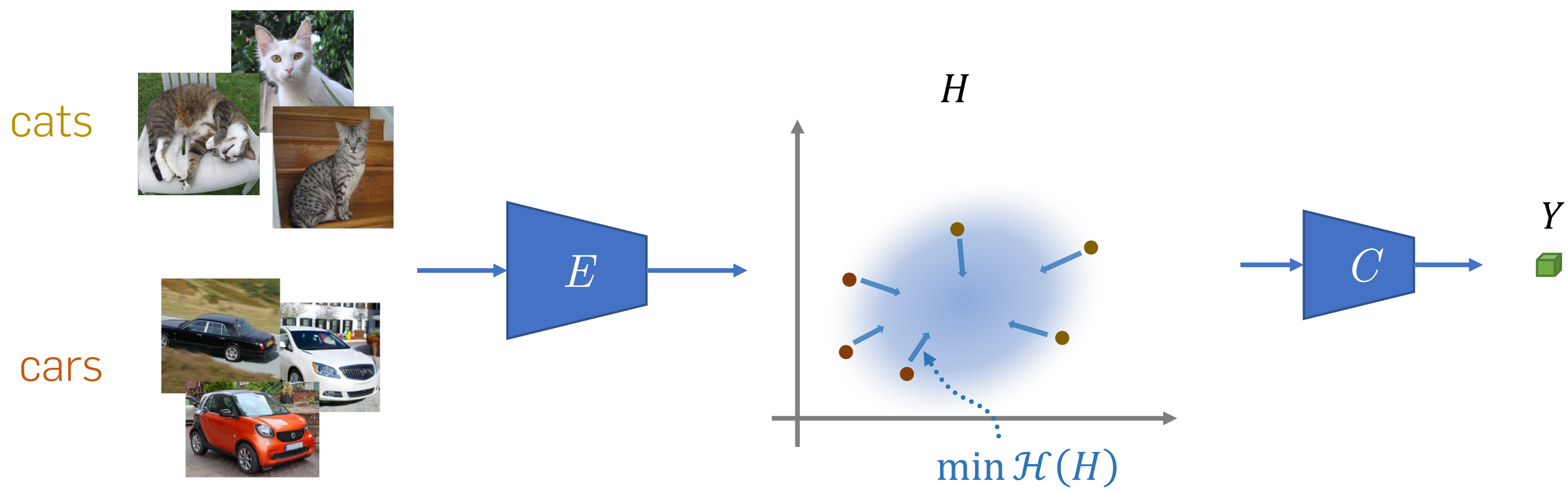
Classification, invariance and entropy

SCIENCES SORBONNE UNIVERSITÉ

$$\min_{\mathbf{w}} \mathcal{L}_{\text{classif}}(\hat{\mathbf{y}}, \mathbf{y})$$

small entropy (noted $\mathcal{H}$) = high invariance

$\mathcal{H}(H)$ ↙ Information in H



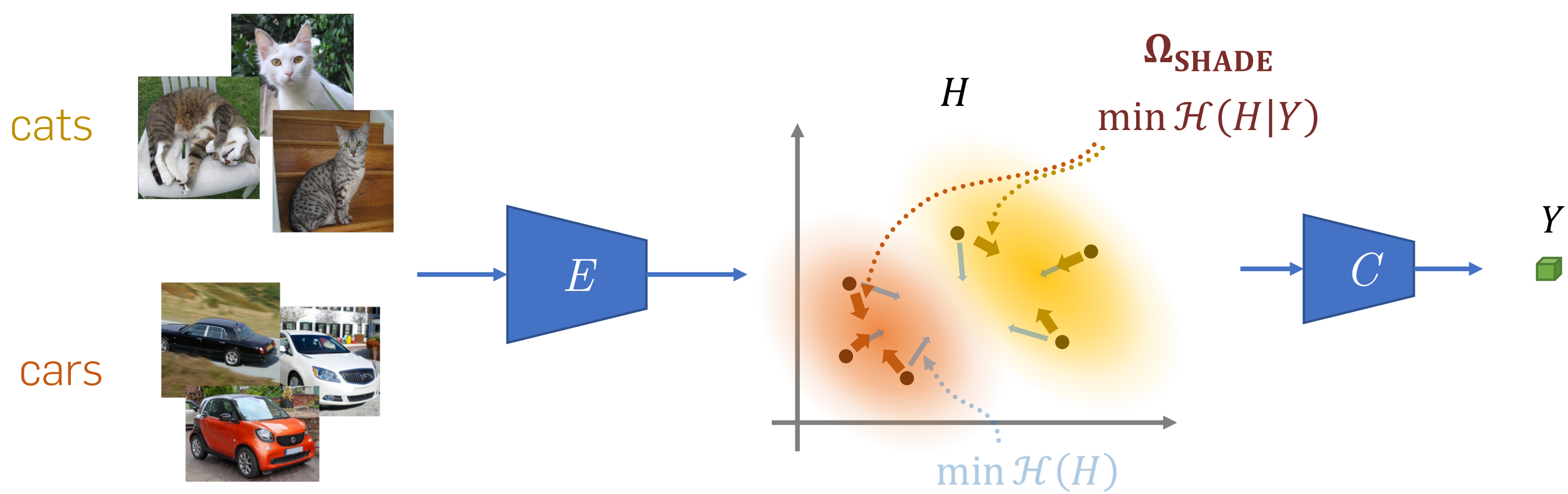
Classification, invariance and entropy

$$\min_{\mathbf{w}} \mathcal{L}_{\text{classif}}(\hat{\mathbf{y}}, \mathbf{y}) + \Omega_{\text{SHADE}}$$

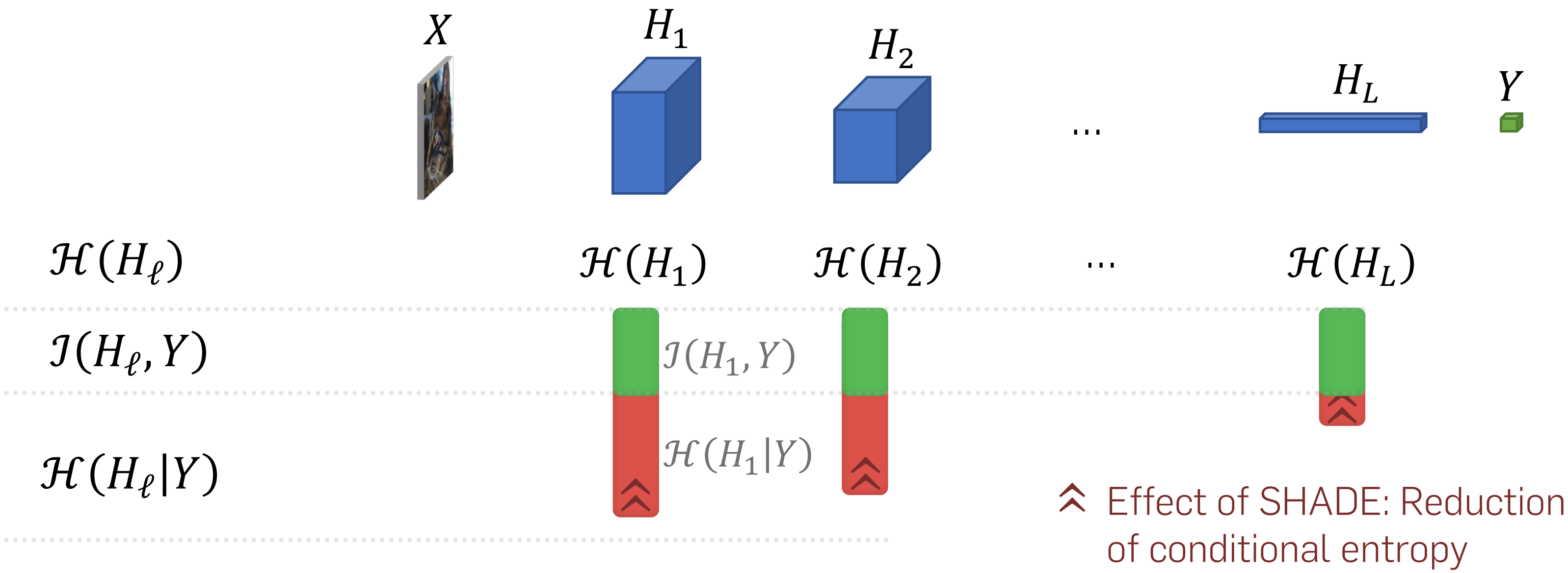
small entropy (noted $\mathcal{H}$) = high invariance

$$\mathcal{H}(H) = \mathcal{I}(H, Y) + \mathcal{H}(H|Y)$$

- $\mathcal{H}(H)$: Information in H
- $\mathcal{I}(H, Y)$: Class-related information
- $\mathcal{H}(H|Y)$: Intra-class variability



SHADE formulation and challenges



$$\Omega_{\text{SHADE}} = \sum_\ell \sum_i \mathcal{H}\left(H_{\ell,i} \middle| Y\right)$$

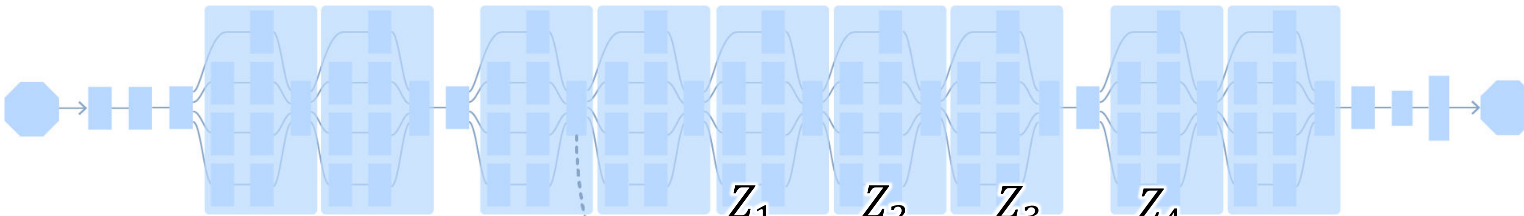
$\mathcal{H}\left(H_{\ell,i} \middle| Y\right)$ is intractable

- → Requires N_{class} entropies $\mathcal{H}\left(H_{\ell,i} \middle| Y = y_k\right) \Rightarrow$ few samples per entropy
- → Complex estimation of entropies

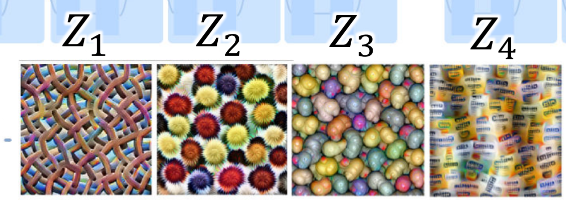
Workarounds – Binary model and variance approx.

SCIENCES SORBONNE UNIVERSITÉ

- Each neuron acts as a binary detector of a specific pattern



Examples of patterns represented:



- We model this detection with a binomial var. Z

$$\begin{cases} p(Z = 1|H) = \text{sigmoid}(H) \\ p(Z = 0|H) = 1 - \text{sigmoid}(H) \end{cases}$$

- Z contains all the information of H useful to predict Y

$$\Omega_{\text{SHADE}} = \mathcal{H}(H \mid Y) = \mathcal{H}(H \mid Z) = \sum_{z \in \{0,1\}} p(Z = z \mid H)\, \mathcal{H}(H \mid Z = z)$$

- Entropy can be approximated by variance

$$\approx \boxed{\sum_{z \in \{0,1\}} p(Z = z \mid H)\, \mathcal{V}\text{ar}(H \mid Z = z)}$$

Training algorithm

SHADE formulation

$$\Omega_{\text{SHADE}} = \sum_{\ell} \sum_{i} \sum_{z \in \{0,1\}} p(Z_{\ell,i} = z | H) \mathcal{V}\text{ar}(H | Z_{\ell,i} = z)$$

$$\boxed{\Omega_{\text{SHADE}} = \sum_{\ell} \sum_{i} \sum_{z \in \{0,1\}} \sum_{k} p\left(Z_{\ell,i} = z \middle| H_{\ell,i}^{(k)}\right) \left(H_{\ell,i}^{(k)} - \mu_{\ell,i}^{z}\right)^2}$$

$\mu_{\ell,i}^{z} \approx \mathbb{E}(H | Z = z)$ with a moving average

Training loss

$$\min_{\mathbf{w}} \mathcal{L}(\mathcal{D}, \mathbf{w}) = \underset{(\mathbf{x},\mathbf{y}) \in \mathcal{D}}{\mathbb{E}} [\mathcal{L}_{\text{classif}}(\hat{\mathbf{y}}, \mathbf{y}) + \Omega_{\text{SHADE}}(H)]$$

Experiments – Comparison to state of the art on CIFAR-10

- → Evaluated for **classification** on **CIFAR-10**
- → Applied on 4 standard deep architectures

Test set accuracy for different regularization methods

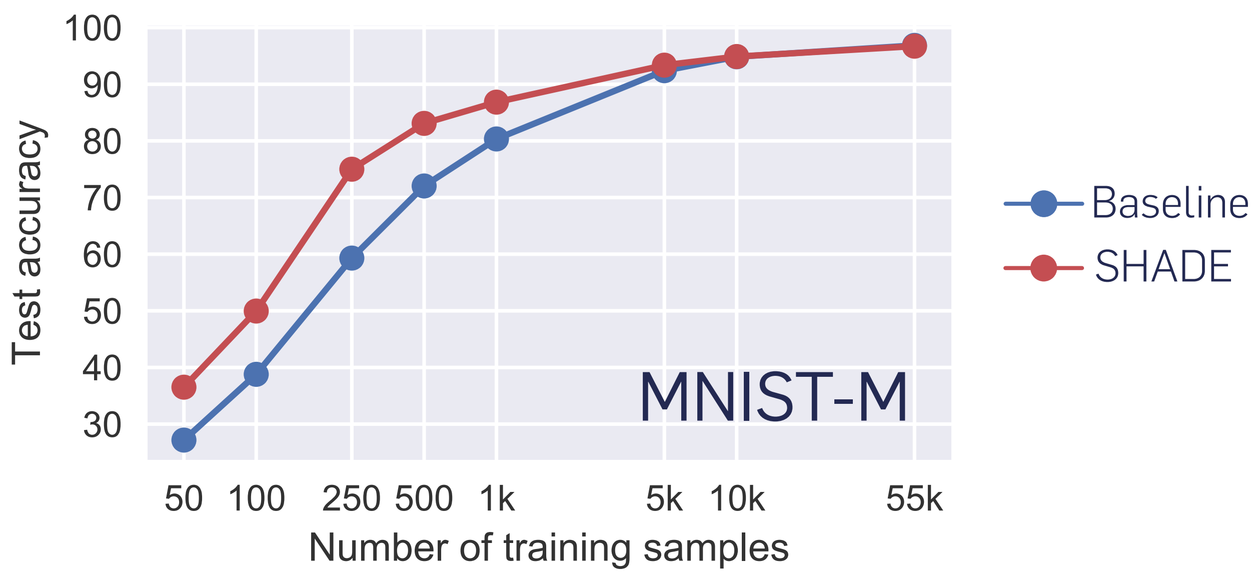
	MLP	AlexNet	ResNet	Inception
No regul.	62.38	83.25	89.84	90.97
Weight decay	62.69	83.54	91.71	91.87
Entropy $\min \mathcal{H}(H)$	63.70	83.61	91.72	91.83
Dropout	65.37	85.95	89.94	91.11
SHADE	66.05	85.45	**92.15**	**93.28**
SHADE + Dropout	**66.12**	**86.71**	92.03	92.51

Experiments – Limited training samples

SCIENCES SORBONNE UNIVERSITÉ

Example of MNIST-M samples

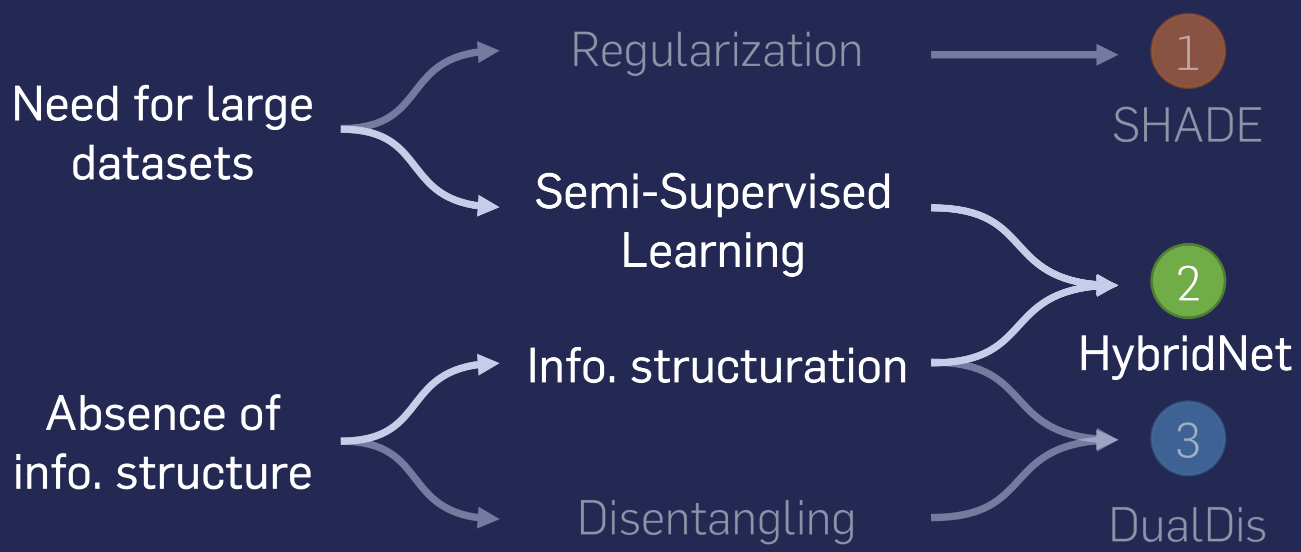
Accuracy with limited train sets



Semi-Supervised Learning

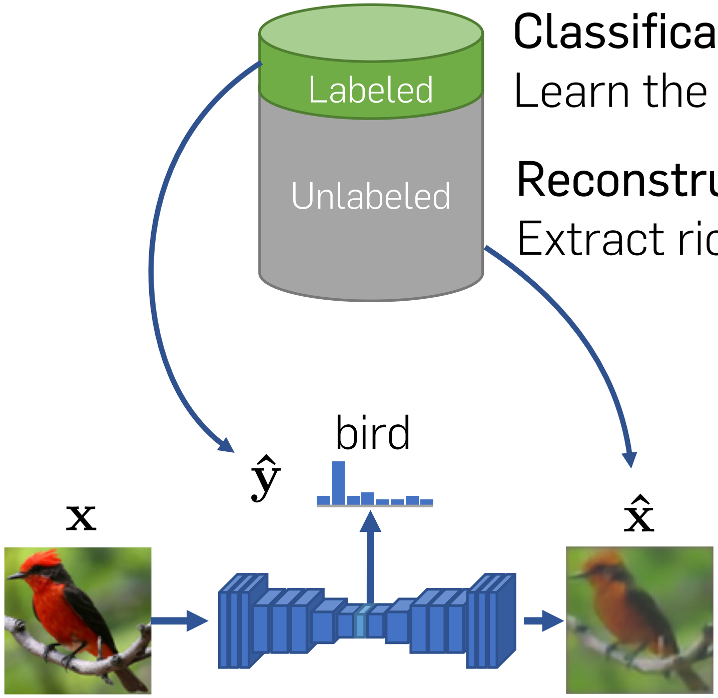
HybridNet: Classification and Reconstruction Cooperation

HybridNet: Classification and Reconstruction Cooperation for Semi-Supervised Learning
Thomas Robert, Nicolas Thome, Matthieu Cord
ECCV 2018



Semi-Supervised Learning

Using unlabeled data to improve the generalization performance

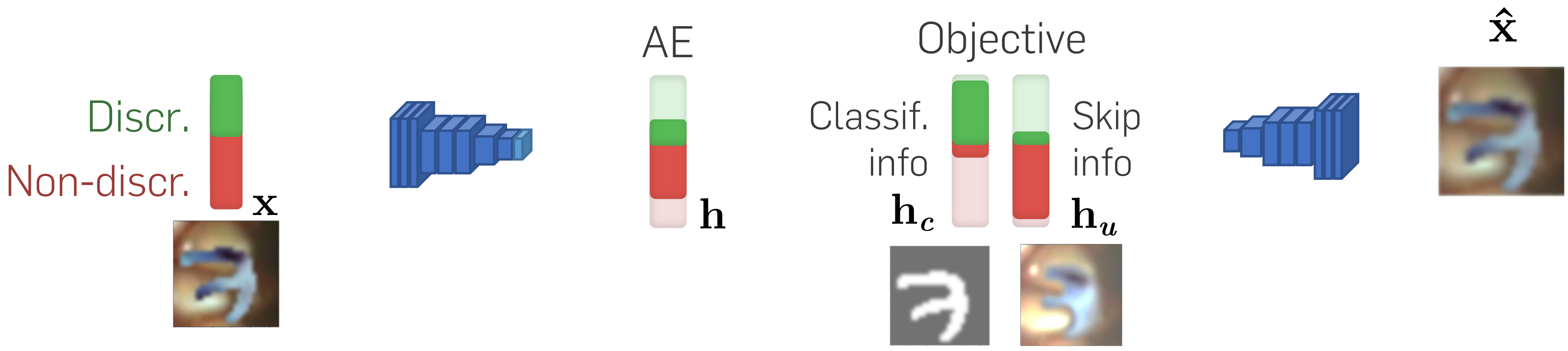


Classification loss
Learn the classification "rule"

Reconstruction loss
Extract rich representative patterns

Motivation and related work

Conflicting goals: Classification → invariance / Reconstruction → info. conservation

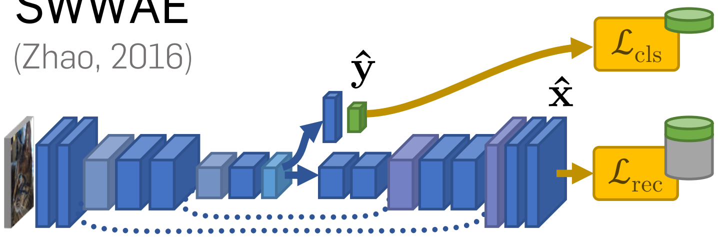


Related work

skip connections for invariance

SWWAE

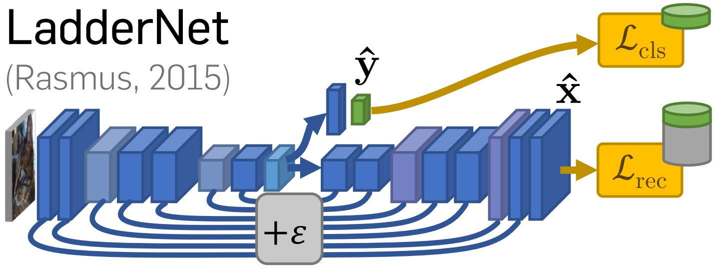
(Zhao, 2016)



Pooling location info

LadderNet

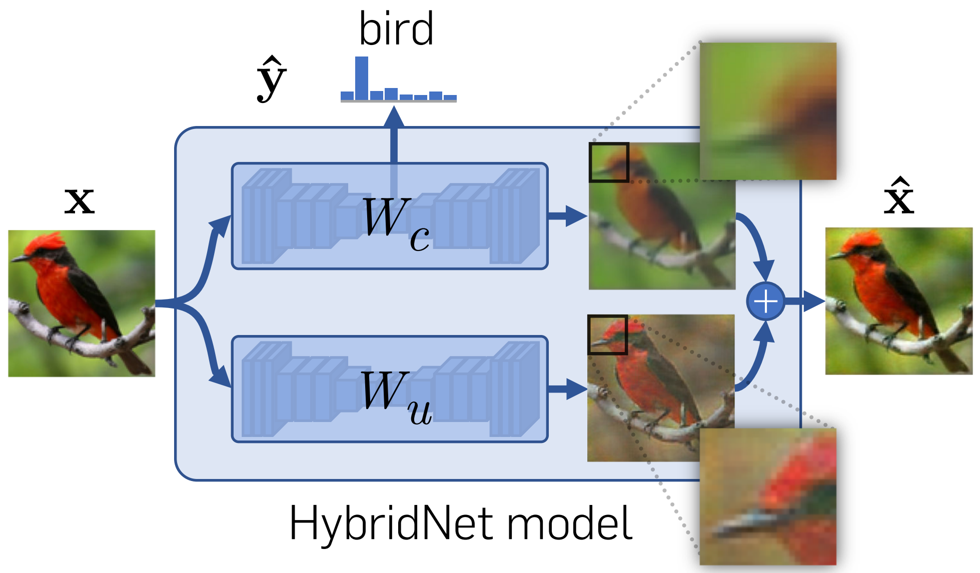
(Rasmus, 2015)



Noisy representations info

Limit: Fixed type of skipped information

HybridNet core proposition



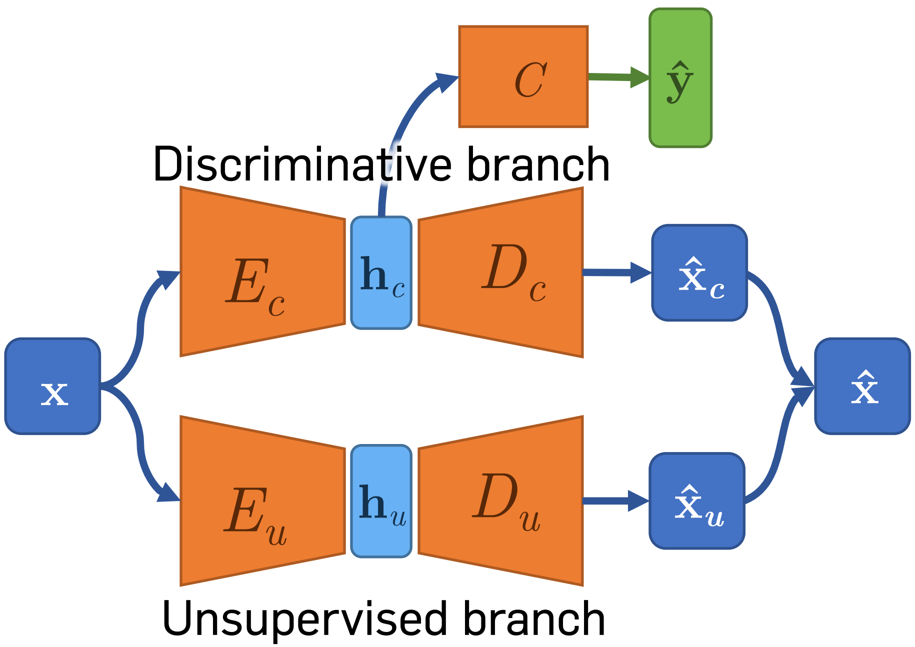
→ **Proposition:** Structure the information in two branches

→ **Goals:**

- → Remove information in the path toward classification
- → Cooperation between the two tasks

Architecture and expected behavior

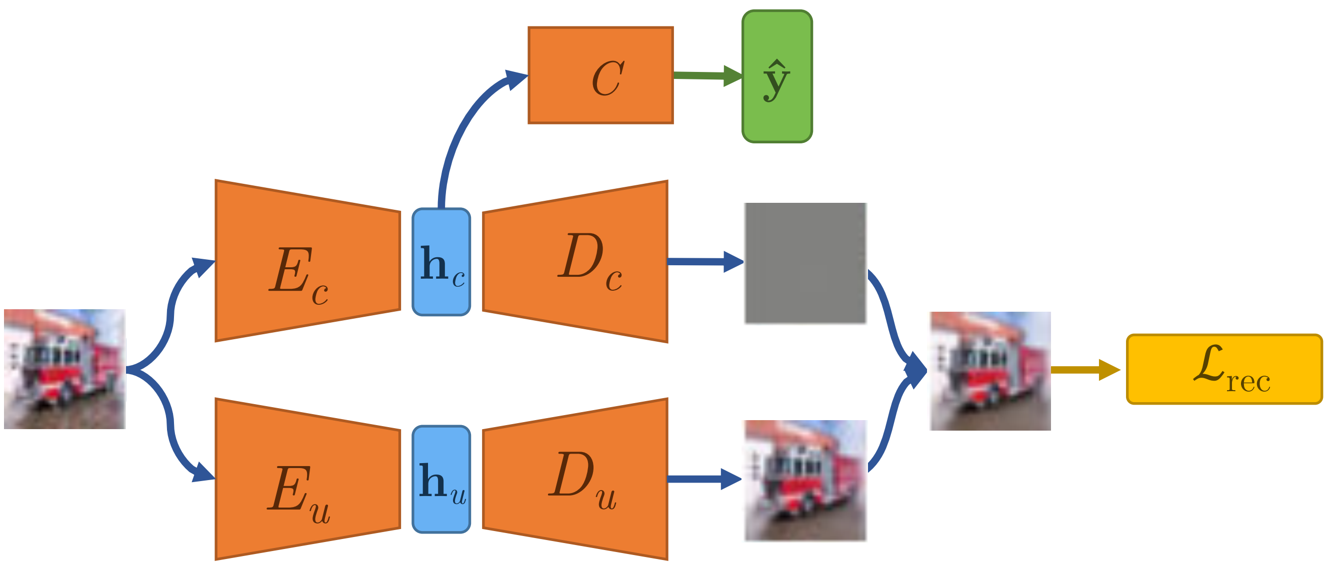
- Discriminative branch
  - Discriminative info only
  - Partial reconstruction
- Unsupervised branch
  - Complementary info & reconstruction
- Challenges:
  - Ensure branch cooperation
  - Guide discriminative features



Training – Branch balancing

SCIENCES SORBONNE UNIVERSITÉ

→ **Problem:** reconstruction from unsup. branch only



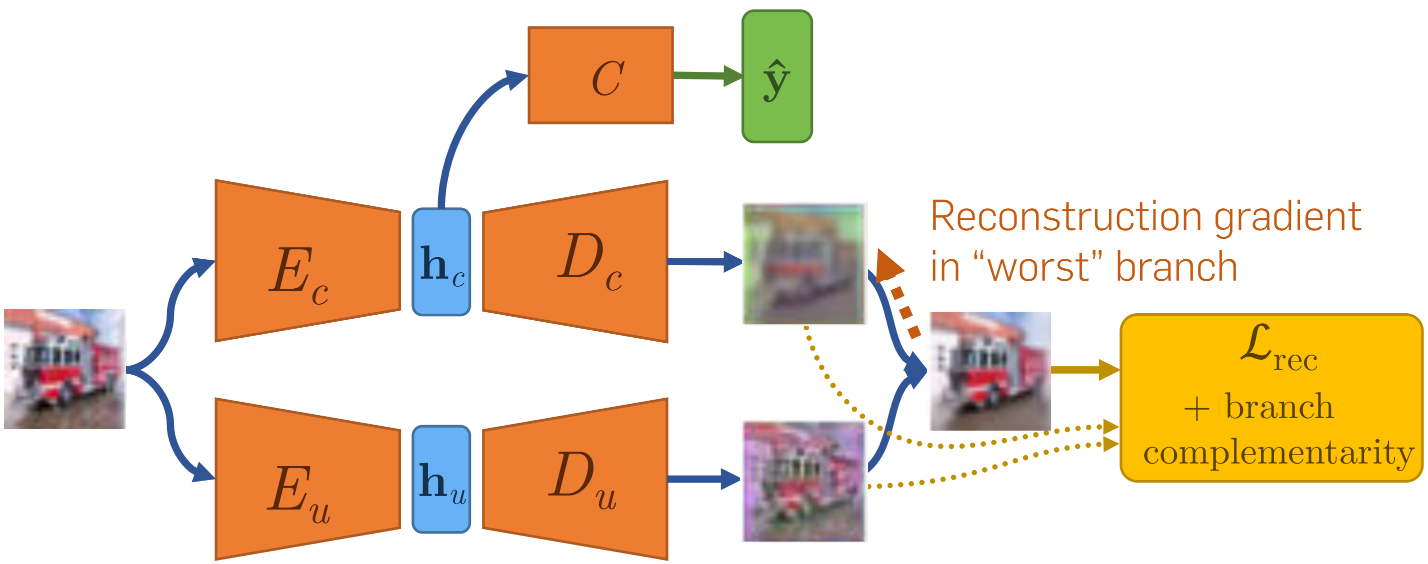
Training – Branch balancing

SCIENCES SORBONNE UNIVERSITÉ

→ **Problem:** reconstruction from unsup. branch only

→ Branch balancing of reconstruction with selective backprop

$$\ell_{\text{br-balance}} = \begin{cases} \|\mathbf{x} - \text{stopgrad}(\hat{\mathbf{x}}_u) - \hat{\mathbf{x}}_c\|_2^2, & \textit{if } \|\mathbf{x} - \hat{\mathbf{x}}_u\| < \|\mathbf{x} - \hat{\mathbf{x}}_c\| \\ \|\mathbf{x} - \hat{\mathbf{x}}_u - \text{stopgrad}(\hat{\mathbf{x}}_c)\|_2^2, & \text{otherwise} \end{cases}$$

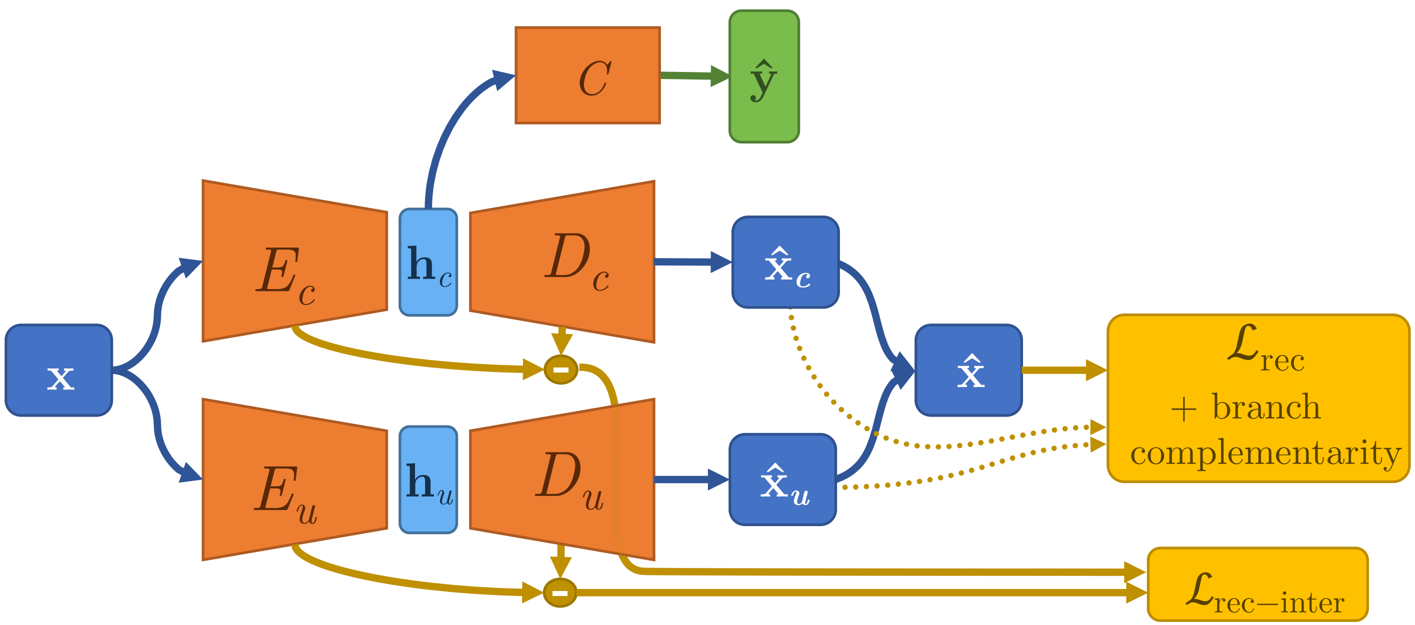


Training – Branch balancing

- **Problem:** reconstruction from unsup. branch only
- Branch balancing of reconstruction with selective backprop

$$\ell_{\text{br-balance}} = \begin{cases} \|\mathbf{x} - \text{stopgrad}(\hat{\mathbf{x}}_u) - \hat{\mathbf{x}}_c\|_2^2, & \textit{if } \|\mathbf{x} - \hat{\mathbf{x}}_u\| < \|\mathbf{x} - \hat{\mathbf{x}}_c\| \\ \|\mathbf{x} - \hat{\mathbf{x}}_u - \text{stopgrad}(\hat{\mathbf{x}}_c)\|_2^2, & \text{otherwise} \end{cases}$$

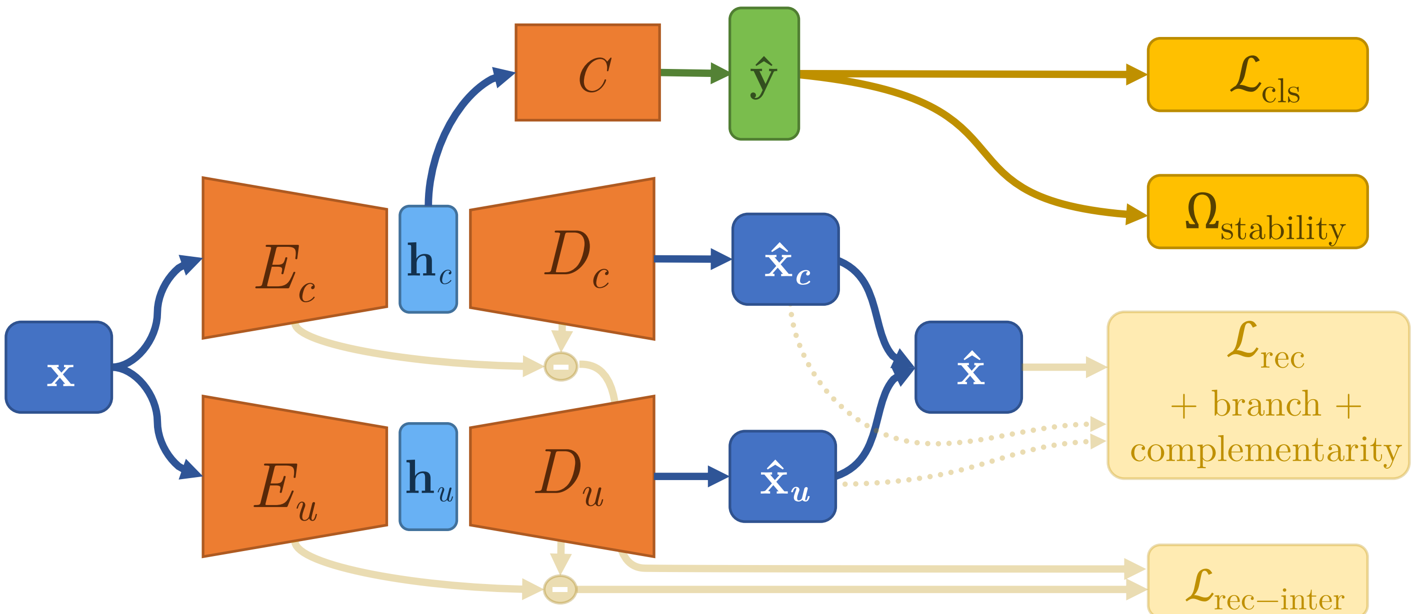
- Intermediate reconstructions



Training – Guiding discriminative features

- Classification loss
- Stability loss e.g. Mean Teacher (Tarvainen, 2017)
  - Encourage invariance to random variability
  - Uses virtual labels $\mathbf{z}^{(i)}$, e.g. avg of outputs

$$\Omega_{\text{stability}} = \left\|\hat{\mathbf{y}}^{(i)} - \mathbf{z}^{(i)}\right\|_2^2$$



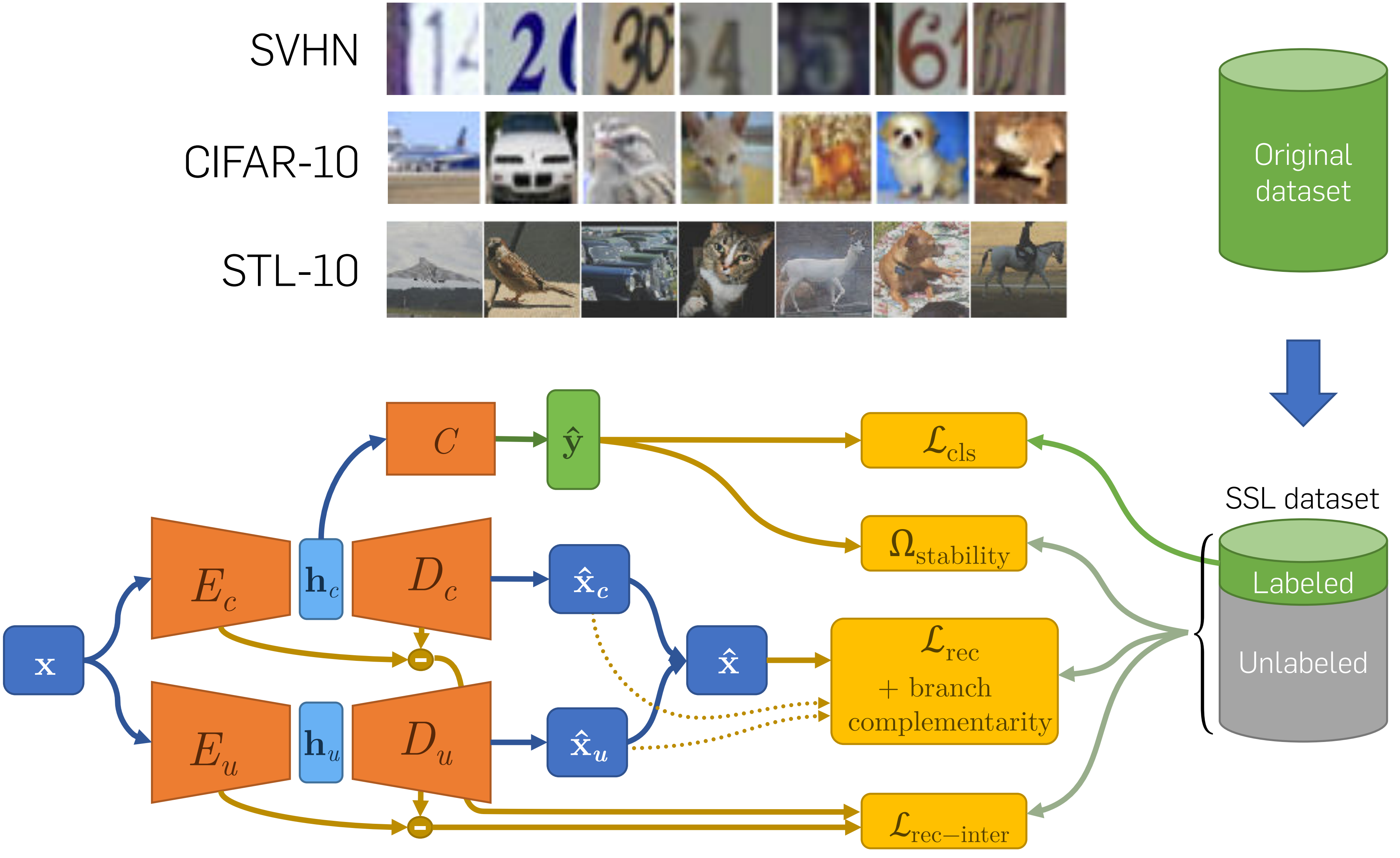
Experiments – Datasets and training

SVHN

CIFAR-10

STL-10

Experiments – Datasets and training

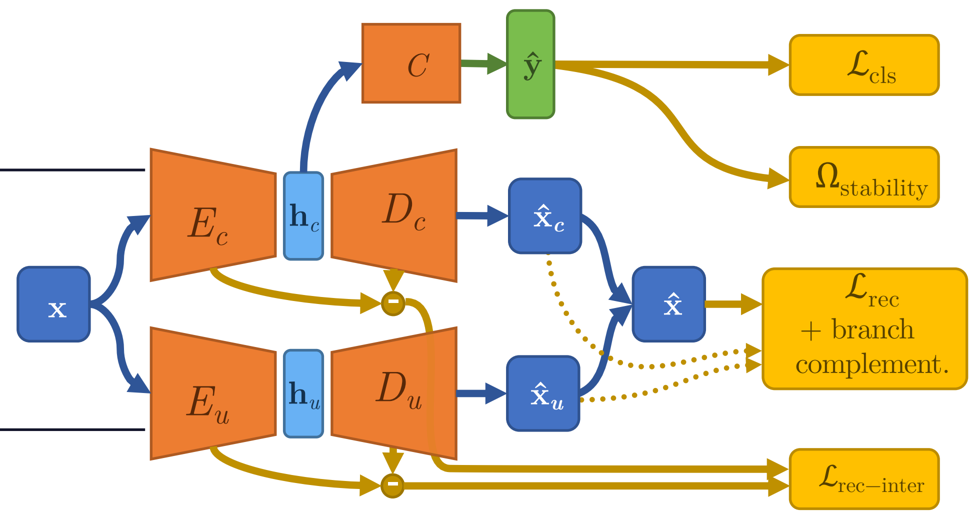


Experiments – Ablation study, quantitative results

SCIENCES SORBONNE UNIVERSITÉ

Accuracy (%) of the model on 2 standard datasets depending on the loss terms used

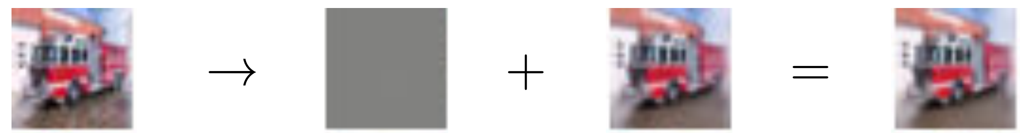
	Classif.	Stability	Reconstruction	Rec. intermed.	Branch balancing	CIFAR-10 2k labels / 50k imgs	STL-10 1k labels / 100k imgs
Classif. baselines	✓					71.5	65.6
	✓	✓				74.6	69.8
HybridNet w/o stability	✓		✓			72.4	67.8
	✓		✓	✓		74.0	
	✓		✓	✓	✓	75.2	
HybridNet w/ stability	✓	✓	✓			77.7	71.5
	✓	✓	✓	✓		80.8	72.2
	✓	✓	✓	✓	✓	**81.6**	**74.1**



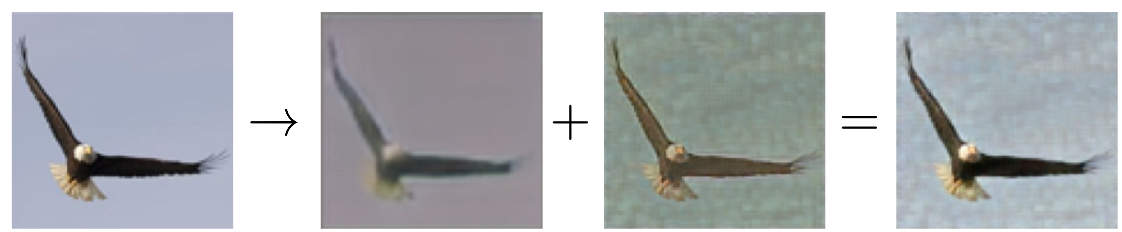
Experiments – Visual analysis

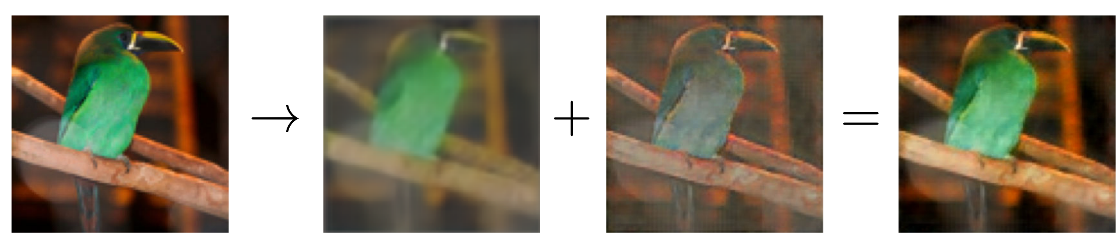
SCIENCES SORBONNE UNIVERSITÉ

Ablation on CIFAR-10

Reconstruction	Rec. intermed.	Branch balacing	Accuracy	Input $\mathbf{x}$ → Discr. branch $\hat{\mathbf{x}}_c$ + Unsup. branch $\hat{\mathbf{x}}_u$ = Rec. $\hat{\mathbf{x}}$
✓			72.4	
✓	✓		74.0	
✓	✓	✓	75.2	

Visualizations on STL-10





Experiments – State-of-the-art results

Error rate (%) on CIFAR-10 test set

	Nb. of labeled imgs	1000	2000	4000
	Ladder network			20.40
	SWWAE (ours impl.)			20.20
	Stability regularization			11.29
	Temporal Ensembling			12.16
	Mean Teacher ConvLarge	21.55	15.73	12.31
ResNet	Supervised baseline	45.22	24.31	15.45
ResNet	Mean Teacher	10.10		6.23
ResNet	**HybridNet**	**8.81**	**7.87**	**6.09**

Notes:

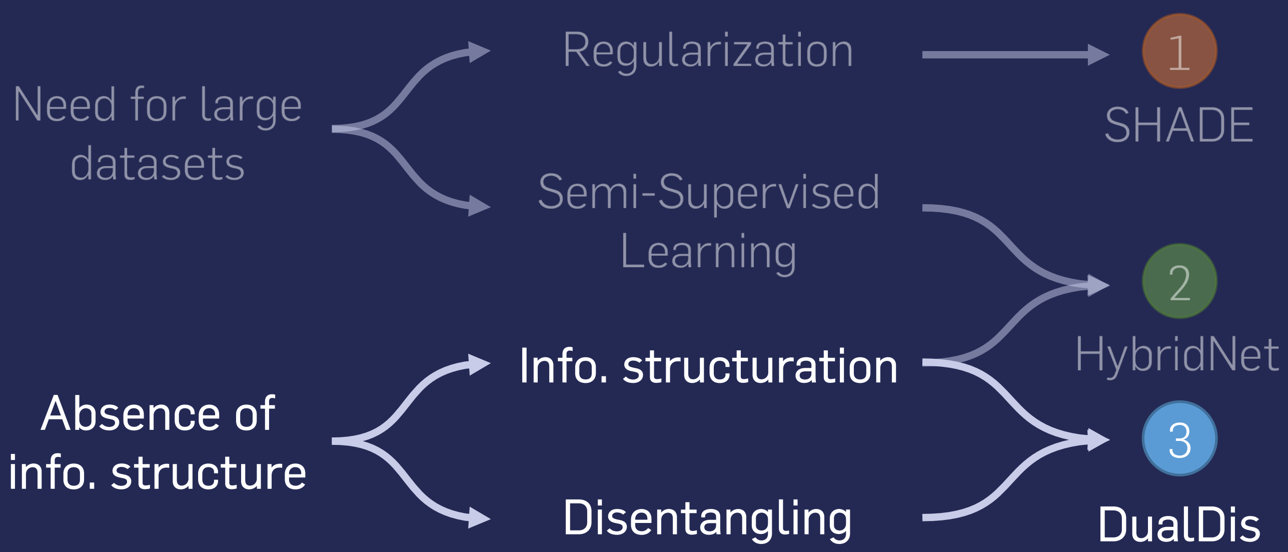
→ SSL comparison must be carried carefully (cf. Oliver, 2018)

→ Similar results on STL-10 and SVHN

Disentangling

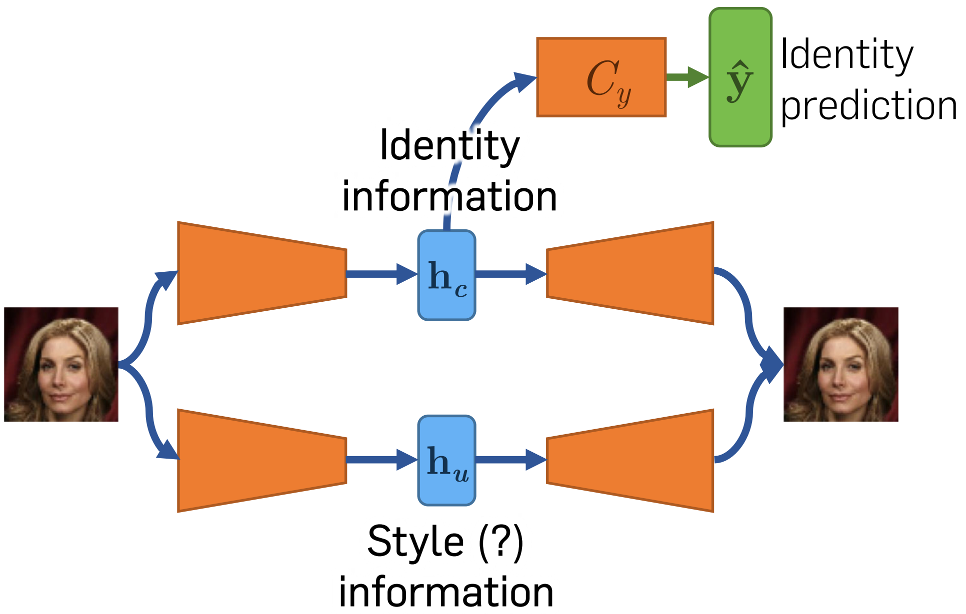
DualDis: Information Separation with Adversarial Learning

DualDis: Dual-Branch Disentangling with Adversarial Learning
Thomas Robert, Nicolas Thome, Matthieu Cord
Under review at AAAI 2020



Toward disentangling and editing

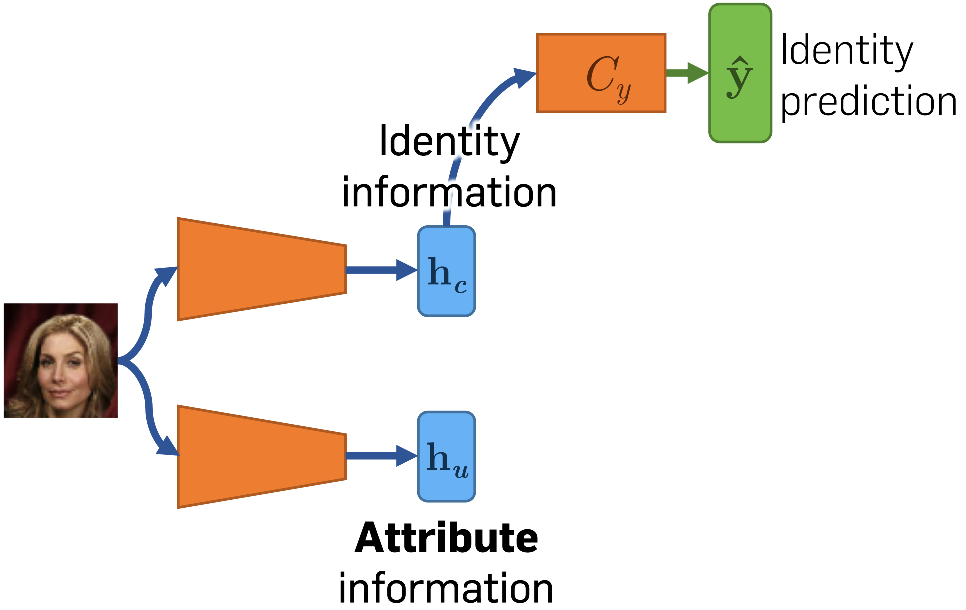
HybridNet



Desired direction: Stronger semantics and information separation between the branches

Toward disentangling and editing

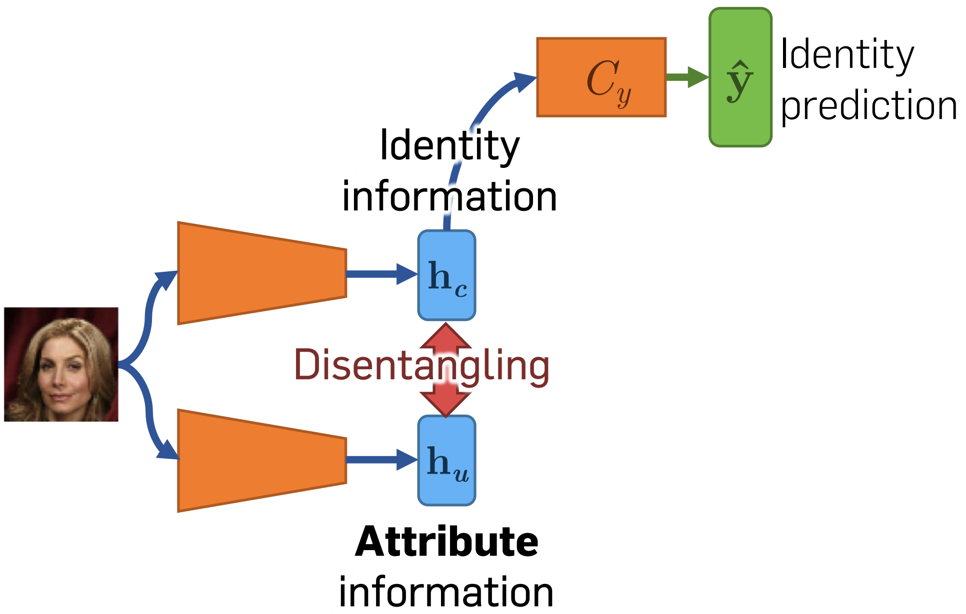
DualDis



- Objectives:
  - Semantic role of each branch

Toward disentangling and editing

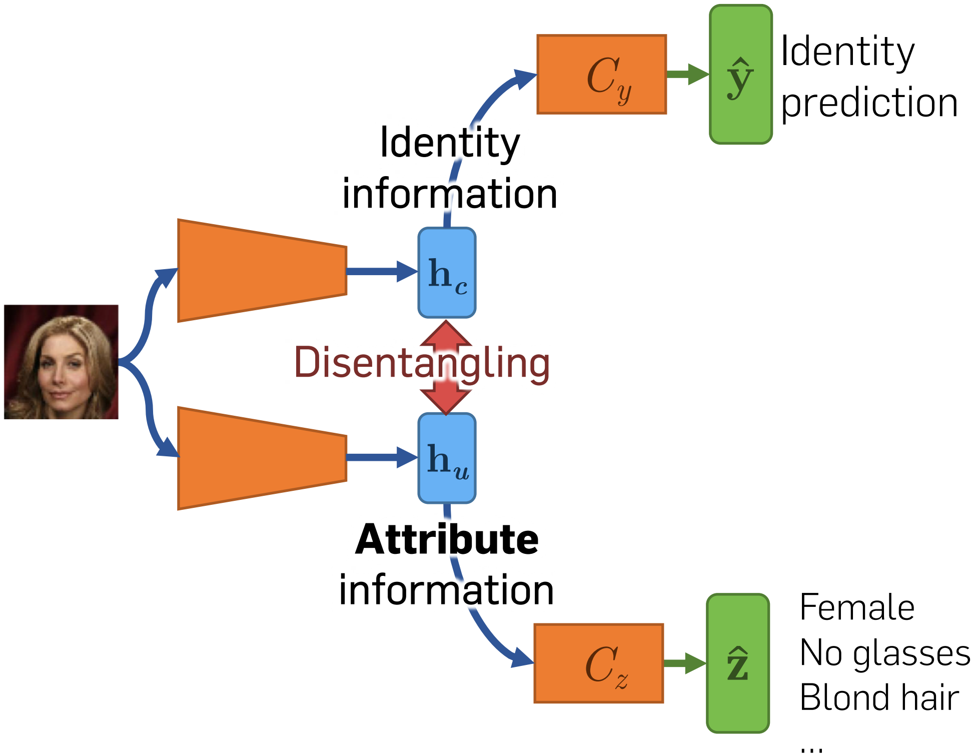
DualDis



- Objectives:
  - **Semantic role** of each branch
  - **Disentangle** (i.e. separate) two information domains

Toward disentangling and editing

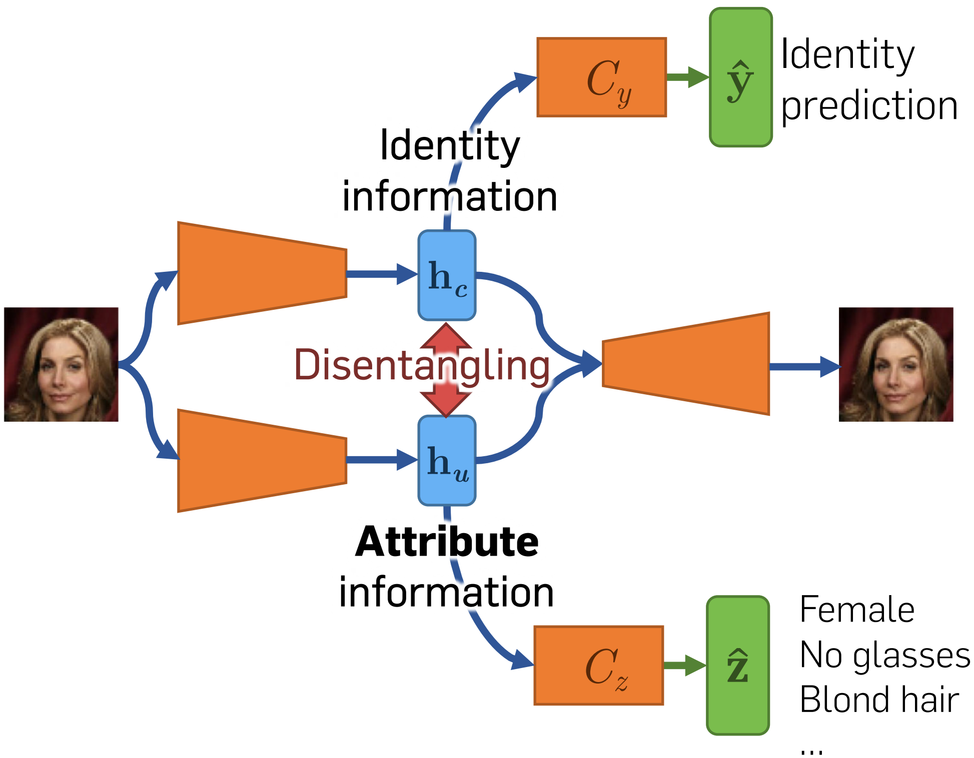
DualDis



- Objectives:
  - **Semantic role** of each branch
  - **Disentangle** (i.e. separate) two information domains
  - **Latent structure** of **semantic factors**

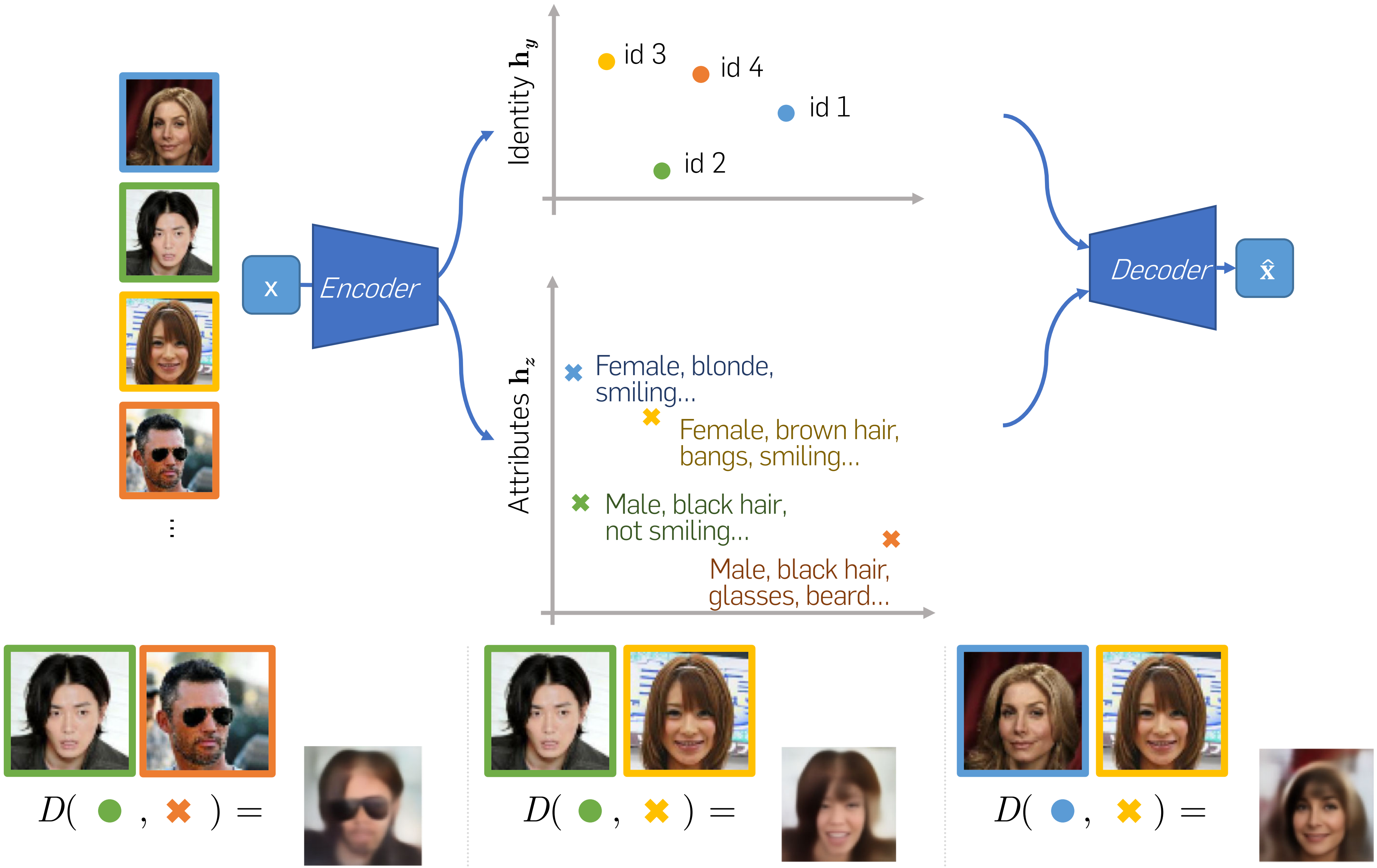
Toward disentangling and editing

DualDis

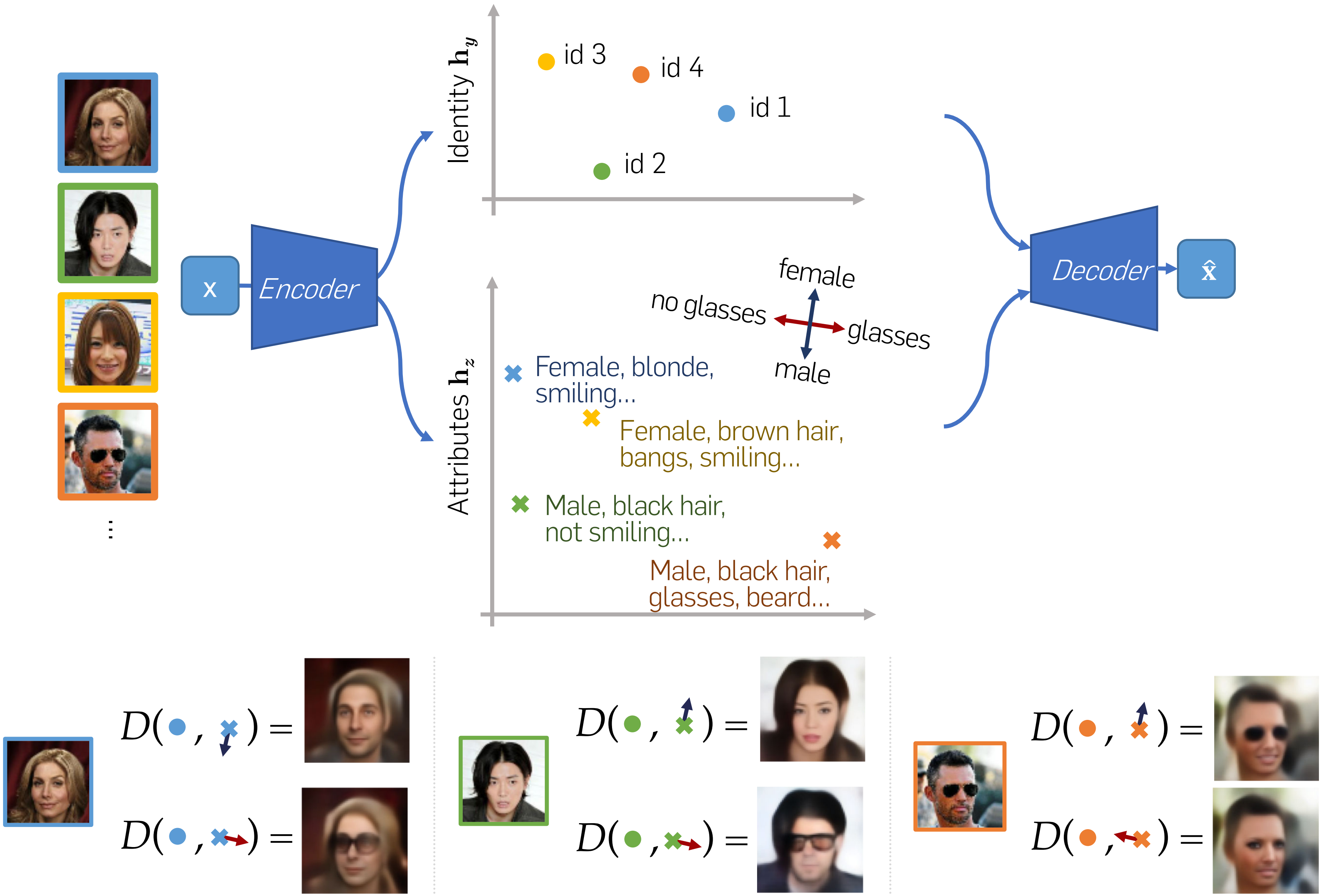


- Objectives:
  - **Semantic role** of each branch
  - **Disentangle** (i.e. separate) two information domains
  - **Latent structure** of **semantic factors**
  - **Image editing** capability

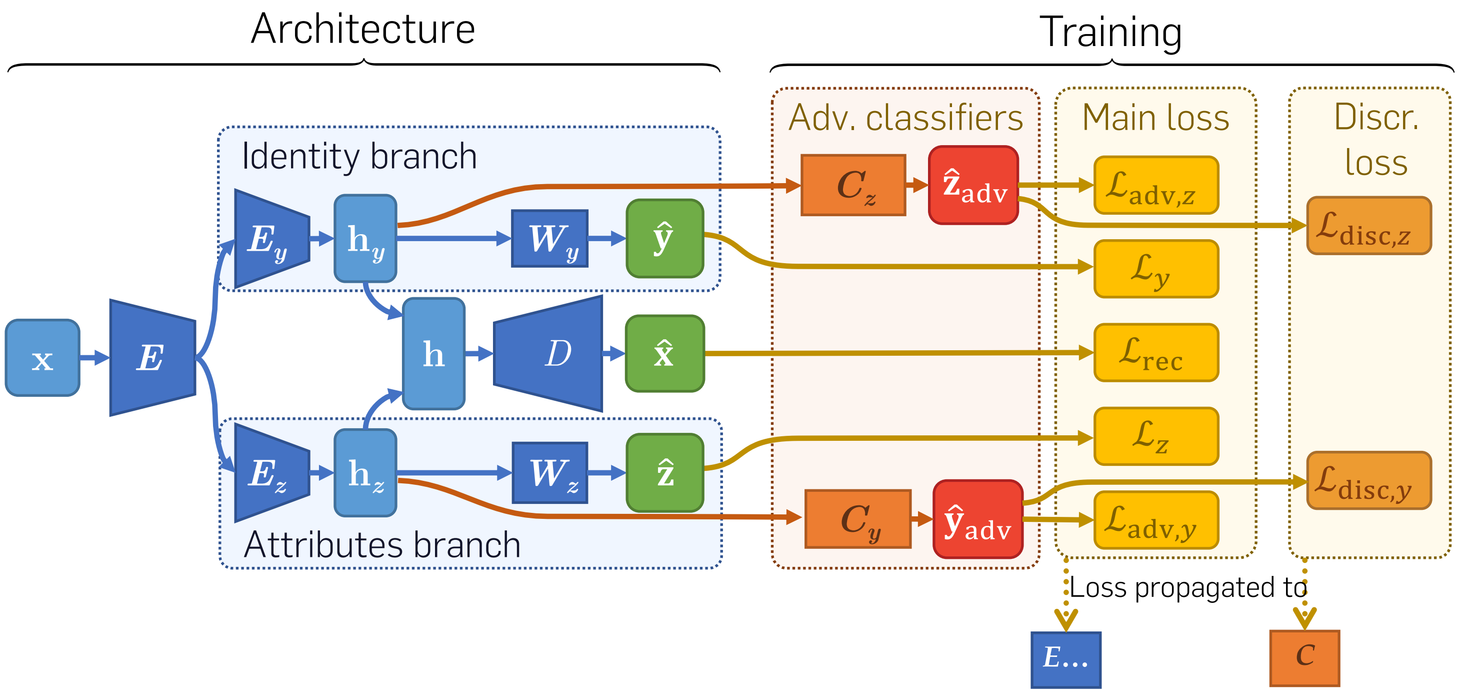
Idea and objectives – Disentangling of domains



Idea and objectives – Intra-domain structure

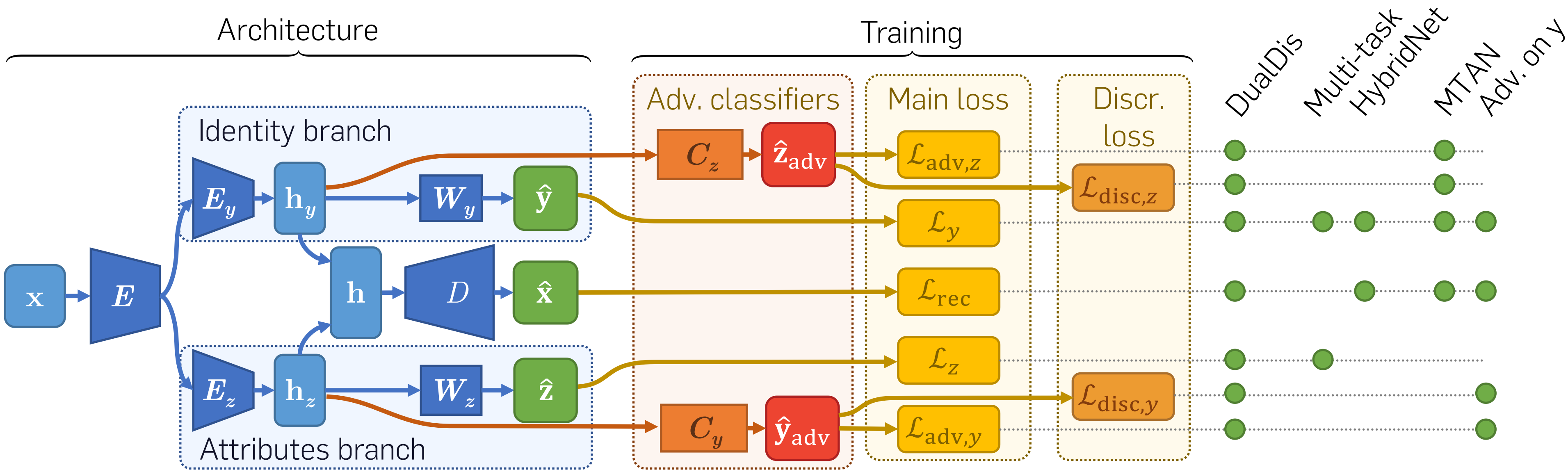


Architecture and training



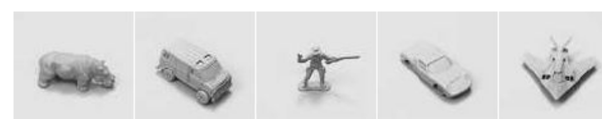
- Two-branch auto-encoder
- Factors representation and linearization by classification
- Disentangling by adversarial classification
  - **Discr. loss:** model undesired info.
  - **Main adv loss:** remove undesired info.

Baselines generalization

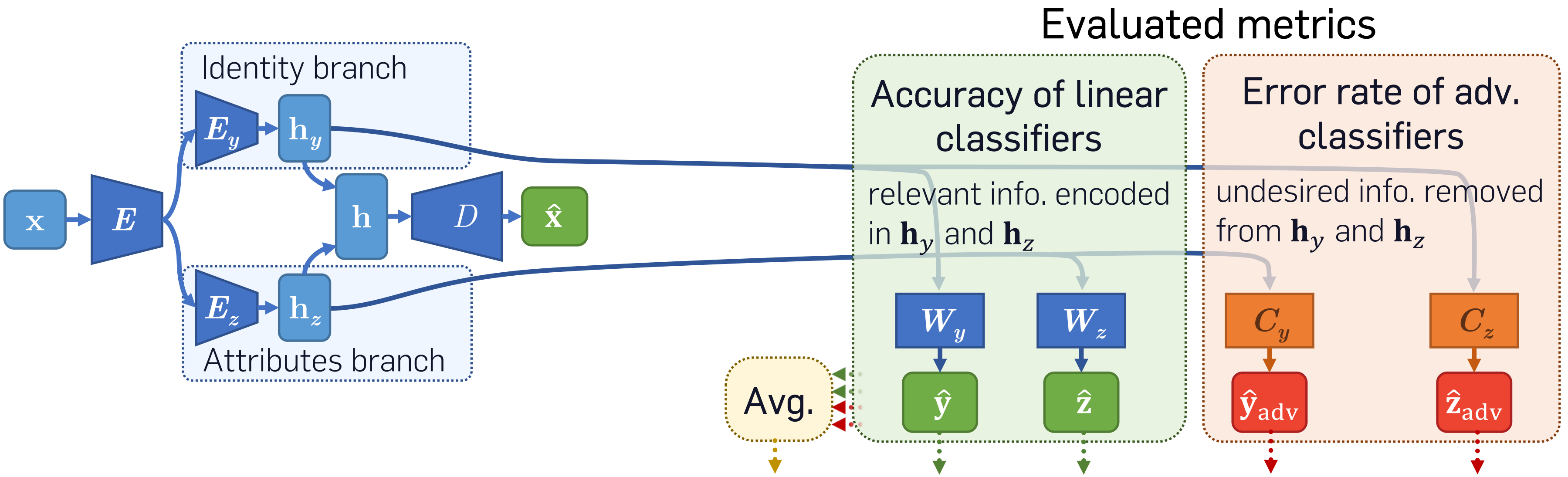


DualDis generalizes many baselines → fair models comparison

Experiments – Datasets

Dataset	Size	Classes (#)	Attr. (#)	Samples
CelebA	60k	Identity (2000)	Style (40)	
Yale-B	2.4k	Identity (38)	Lighting (14)	
NORB	48k	Category (5)	Lighting + pose (8)	

Experiments – Quantitative results (Yale-B)



Model	Labels used	Aggr. Metric	Accuracy $\mathbf{h}_y \to \mathbf{y}$	Accuracy $\mathbf{h}_z \to \mathbf{z}$	Disentangling $\mathbf{h}_z \to \mathbf{y}_{adv}$	Disentangling $\mathbf{h}_y \to \mathbf{z}_{adv}$
Multi-task classif.	$\mathbf{y}, \mathbf{z}$	81.5	98.5%	**97.2%**	85.3%	45.1%
HybridNet-like	$\mathbf{y}$	65.3	97.6%	93.7%	23.3%	46.5%
HybridNet-like + attr	$\mathbf{y}, \mathbf{z}$	80.5	**99.0%**	96.9%	80.0%	46.1%
MTAN (Liu, 2018)	$\mathbf{y}, \mathbf{z}, \mathbf{z}_{test}$	—	98.4%	—	—	70.3%
Adv. on y only (Hadad, 2018)	$\mathbf{y}$	79.8	98.3%	84.1%	92.5%	44.4%
DualDis	$\mathbf{y}, \mathbf{z}$	**92.0**	98.6%	**97.3%**	**98.8%**	**73.4%**

Experiments – Semi-supervised learning

- **Limit of DualDis:** requires attributes labels
- Need can be reduced by SSL

SSL results on CelebA

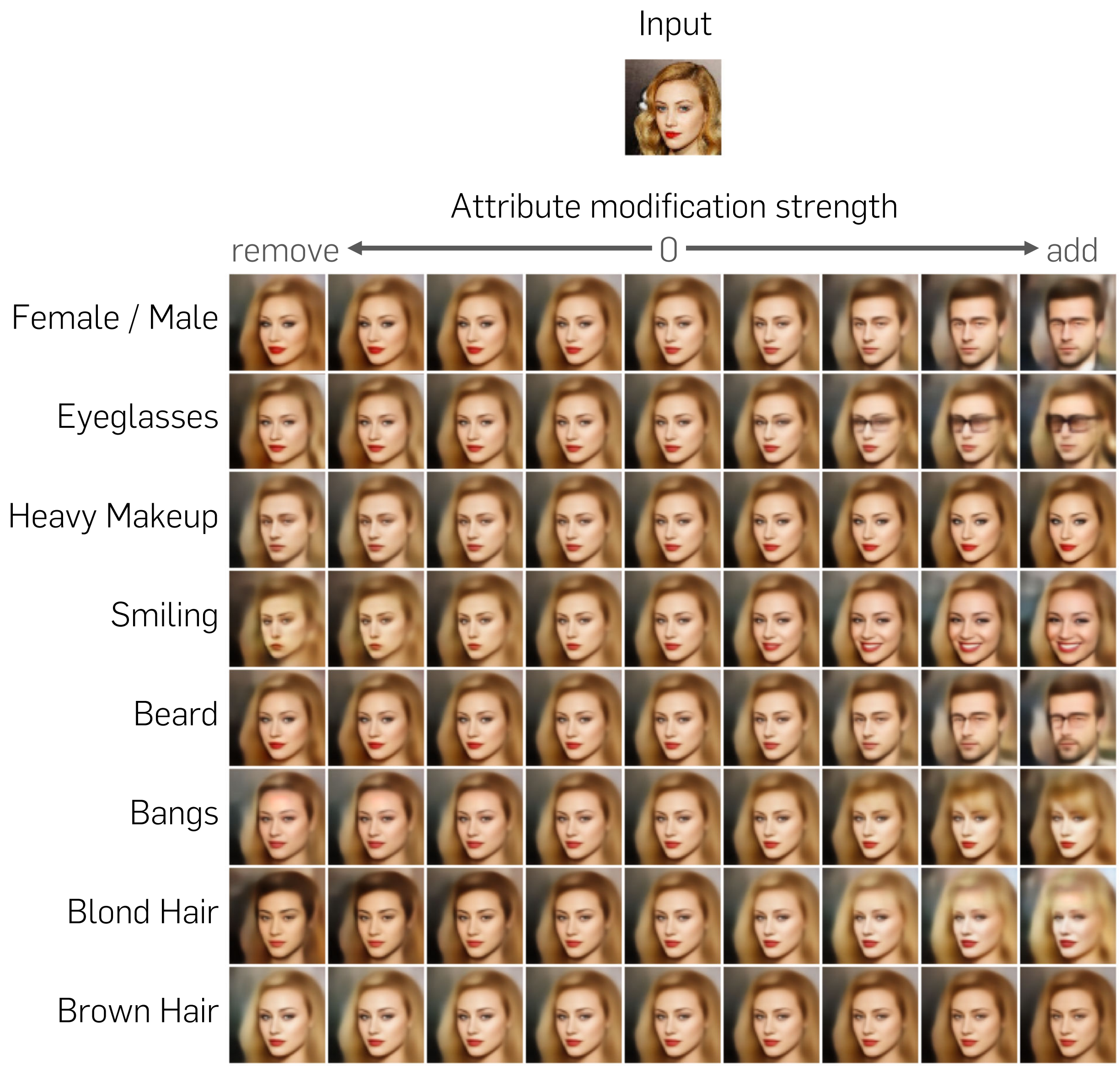
Nb. attr. labels	Aggr. metric	Accuracy		Disentangling	
		$\mathbf{h}_y \rightarrow \mathbf{y}$	$\mathbf{h}_z \rightarrow \mathbf{z}$	$\mathbf{h}_z \rightarrow \mathbf{y}_{\text{adv}}$	$\mathbf{h}_y \rightarrow \mathbf{z}_{\text{adv}}$
2000	66.8	71.0%	85.0%	**98.4%**	12.7%
48000 (full labels)	68.0	71.1%	**88.6%**	97.3%	**14.9%**

Experiments – Image editing

Represent an image
$E(\mathbf{x}) \rightarrow \mathbf{h}_y, \mathbf{h}_z$

Edit the representation
$\mathbf{h}'_z = \mathbf{h}_z + \varepsilon \mathbf{w}_{z,i}$

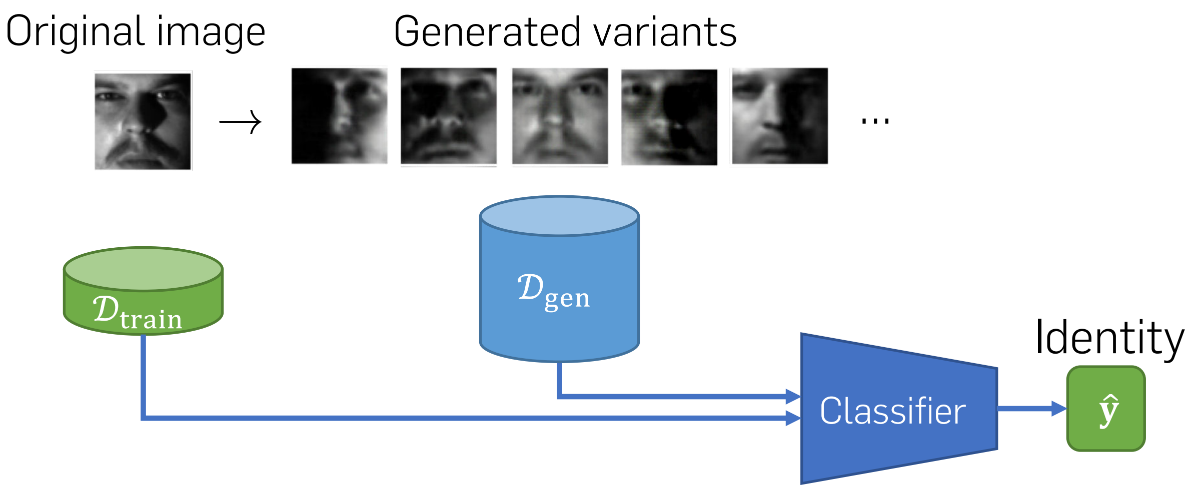
Generate a new image
$D(\mathbf{h}_y, \mathbf{h}'_z) = \hat{\mathbf{x}}'$



Experiments – Semantic data augmentation

SCIENCES SORBONNE UNIVERSITÉ

→ For each train image, generate variations in attributes

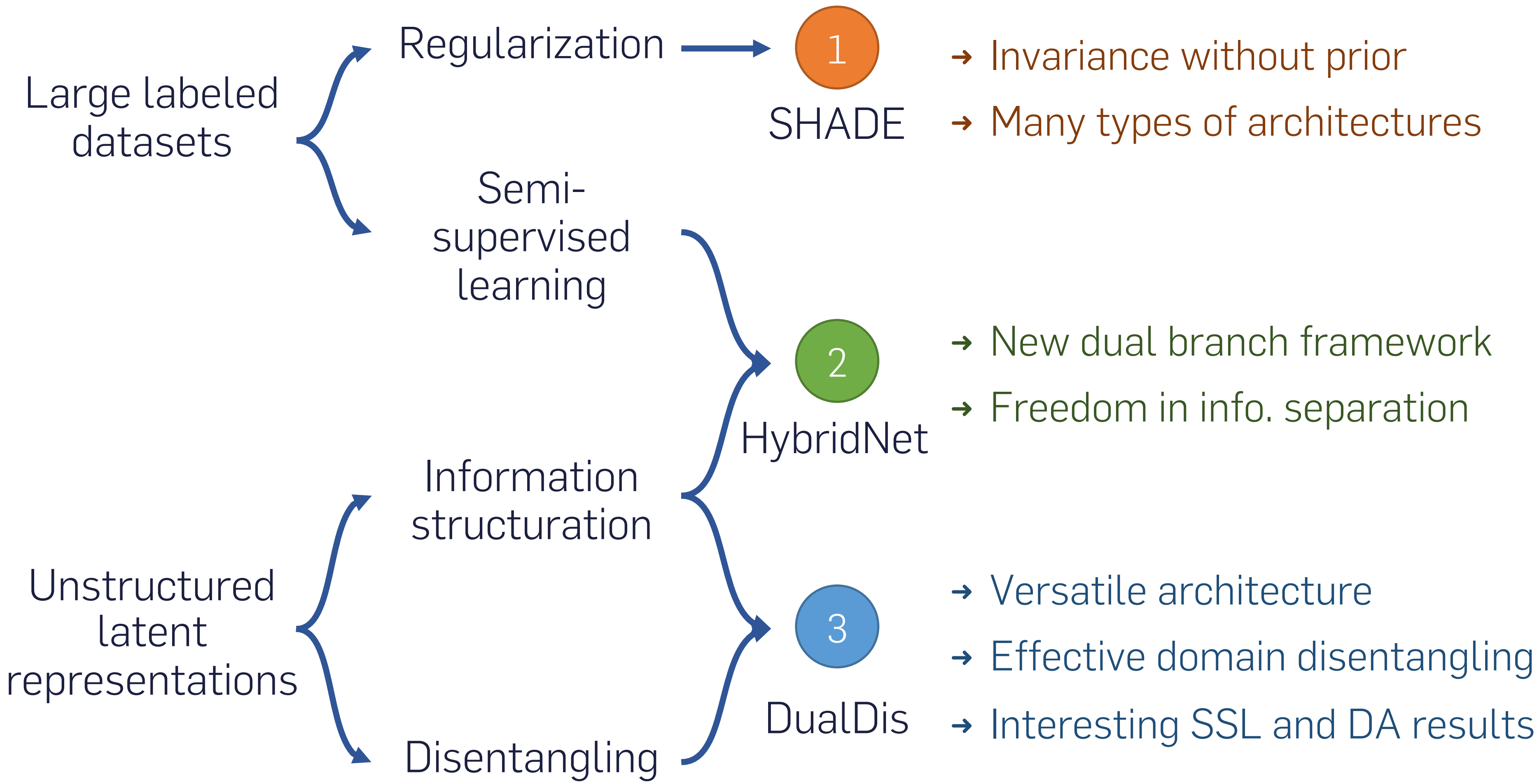


Classifier accuracy on Yale-B test set

$\lvert\mathcal{D}_{\text{train}}\rvert$ \ Train set	$\mathcal{D}_{\text{train}}$ Baseline	$\mathcal{D}_{\text{train}} + \mathcal{D}_{\text{gen}}$ Nb. generated samples per class: 10	20	30	60
240	48.9%	51.8%	55.5%	56.8%	**58.6%**
360	69.1%	70.5%	72.6%	73.1%	**75.6%**
480	78.9%	79.3%	80.1%	81.6%	**82.8%**

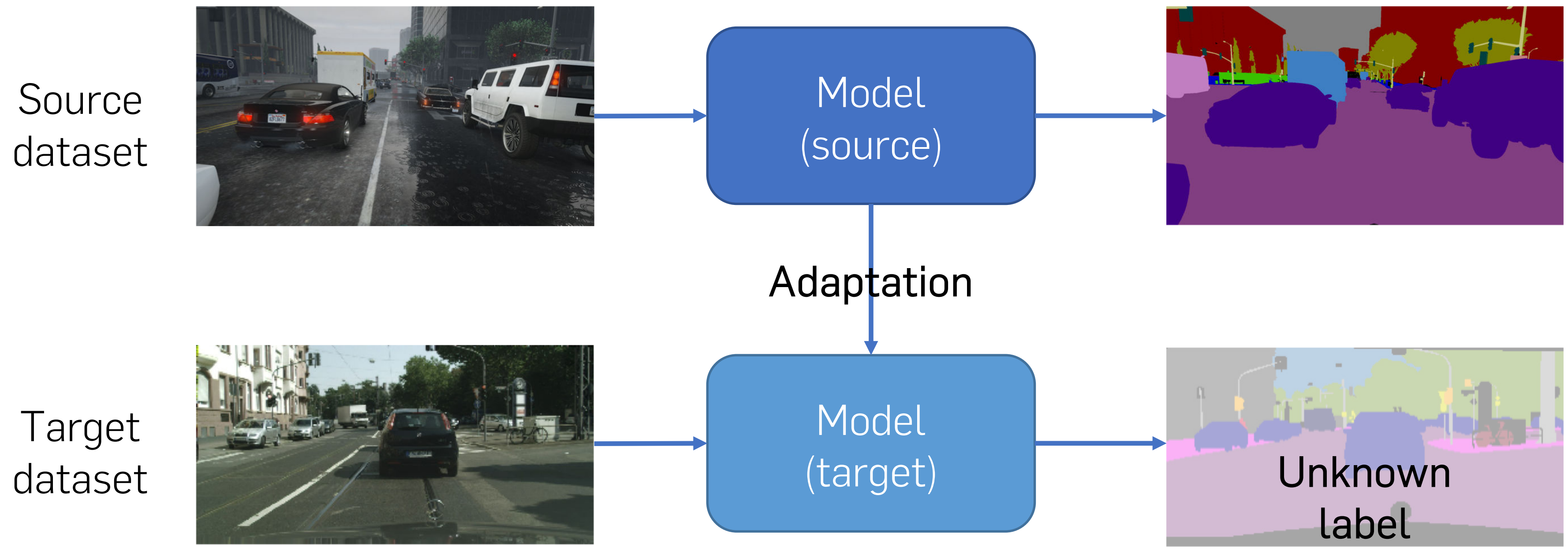
Conclusion

Contributions



Perspectives

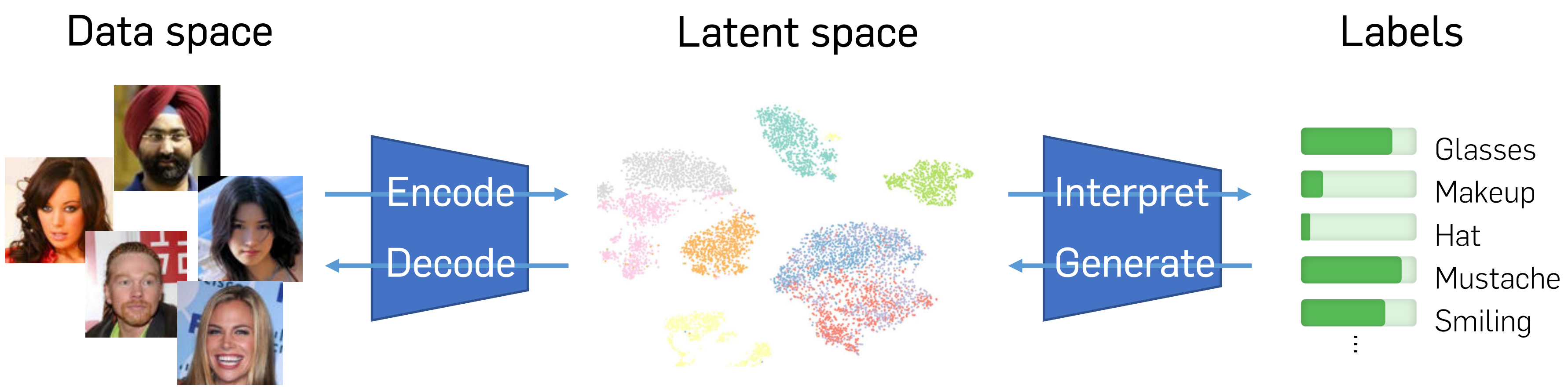
- Domain adaptation
  - Separate domain specific and domain agnostic features
  - Generate data for the target domain



Perspectives

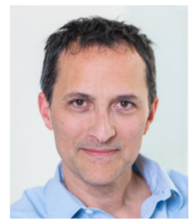
SCIENCES SORBONNE UNIVERSITÉ

- Bridging the gap between discriminative and generative models
  - Reversible models
  - Latent structure & model of factors' internal diversity
  - Semantic data augmentation



Thank You!

Thomas Robert
MLIA



Matthieu Cord
MLIA
Thesis director

Nicolas Thome
CNAM
Thesis co-director

Publications:

DualDis: Dual-Branch Disentangling with Adversarial Learning.
T. Robert, N. Thome, M. Cord. Under review, AAAI 2020.

HybridNet: Classification and Reconstruction Cooperation for Semi-Supervised Learning.
T. Robert, N. Thome, M. Cord. ECCV, 2018.

SHADE: Information-Based Regularization for Deep Learning.
M. Blot, T. Robert, N. Thome, M. Cord. Best paper ICIP, 2018.

Machine Learning & Deep Learning for Information Access

References

•

Appendix

References

- Achille, A. and S. Soatto (2016). “Information Dropout: learning optimal representations through noisy computation”. In: arXiv.
- Alemi, Alexander, Ian Fischer, Joshua V. Dillon, and Kevin Murphy (2017). “Deep Variational Information Bottleneck”. In: ICLR.
- Bousmalis, Konstantinos, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan (2016). “Domain separation networks”. In: NIPS.
- Chang, Wei-Lun, Hui-Po Wang, Wen-Hsiao Peng, and Wei-Chen Chiu (2019). “All about Structure: Adapting Structural Information across Domains for Boosting Semantic Segmentation”. In: CVPR.
- Chen, Tian Qi, Xuechen Li, Roger B Grosse, and David K Duvenaud (2018). “Isolating sources of disentanglement in variational autoencoders”. In: NeurIPS.
- Dupont, Emilien (2018). “Learning disentangled joint continuous and discrete representations”. In: NeurIPS.
- Dyk, David A. van and Xiao-Li Meng (2001). “The Art of Data Augmentation”. In: Journal of Computational and Graphical Statistics.
- Engel, Jesse, Matthew Hoffman, and Adam Roberts (2018). “Latent Constraints: Learning to Generate Conditionally from Unconditional Generative Models”. In: ICLR.
- Gomez, Aidan N, Mengye Ren, Raquel Urtasun, and Roger B Grosse (2017). “The Reversible Residual Network: Backpropagation Without Storing Activations”. In: NIPS.
- Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio (2014). “Generative Adversarial Nets”. In: NIPS.
- Hadad, Naama, Lior Wolf, and Moni Shahar (2018). “A Two-Step Disentanglement Method”. In: CVPR.
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2016). “Deep Residual Learning for Image Recognition”. In: CVPR.

References

- He, Zhenliang, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen (2019). “Attgan: Facial attribute editing by only changing what you want”. In: TIP.
- Higgins, Irina, David Amos, David Pfau, Sebastien Racaniere, Loic Matthey, Danilo Rezende, and Alexander Lerchner (2018). “Towards a Definition of Disentangled Representations”. In: arXiv.
- Higgins, Irina, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner (2017). “beta-vae: Learning basic visual concepts with a constrained variational framework”. In: ICLR.
- Hu, Qiyang, Attila Szabó, Tiziano Portenier, Paolo Favaro, and Matthias Zwicker (2018). “Disentangling Factors of Variation by Mixing Them”. In: CVPR.
- Ioffe, Sergey and Christian Szegedy (2016). “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: JMLR.
- Jacobsen, Jörn-Henrik, Arnold Smeulders, and Edouard Oyallon (2018). “i-RevNet: Deep Invertible Networks”. In: ICLR.
- Jaiswal, Ayush, Rex Yue Wu, Wael Abd-Almageed, and Prem Natarajan (2018). “Unsupervised Adversarial Invariance”. In: NeurIPS.
- Karras, Tero, Samuli Laine, and Timo Aila (2019). “A style-based generator architecture for generative adversarial networks”. In: CVPR.
- Kim, Hyunjik and Andriy Mnih (2018). “Disentangling by factorising”. In: arXiv.
- Kingma, Diederik P and Prafulla Dhariwal (2018). “Glow: Generative flow with invertible 1x1 convolutions”. In: NeurIPS.
- Klys, Jack, Jake Snell, and Richard Zemel (2018). “Learning Latent Subspaces in Variational Autoencoders”. In: NeurIPS.

References

- Kokkinos, Iasonas (2017). "Ubernet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory". In: CVPR.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton (2012). "Imagenet classification with deep convolutional neural networks". In: NIPS.
- Krogh, Anders and John A. Hertz (1992). "A Simple Weight Decay Can Improve Generalization". In: NIPS.
- Kulkarni, Tejas D, William F. Whitney, Pushmeet Kohli, and Josh Tenenbaum (2015). "Deep Convolutional Inverse Graphics Network". In: NIPS.
- Laine, Samuli and Timo Aila (2017). "Temporal Ensembling for Semi-Supervised Learning". In: ICLR.
- Lample, Guillaume, Neil Zeghidour, Nicolas Usunier, Antoine Bordes, Ludovic Denoyer, and Marc'Aurelio Ranzato (2017). "Fader Networks: Manipulating Images by Sliding Attributes". In: NIPS.
- Liu, Yang, Zhaowen Wang, Hailin Jin, and Ian Wassell (2018). "Multi-Task Adversarial Network for Disentangled Feature Learning". In: CVPR.
- Liu, Yu, Fangyin Wei, Jing Shao, Lu Sheng, Junjie Yan, and Xiaogang Wang (2018). "Exploring Disentangled Feature Representation Beyond Face Identification". In: CVPR.
- Liu, Ziwei, Ping Luo, Xiaogang Wang, and Xiaoou Tang (2015). "Deep Learning Face Attributes in the Wild". In: ICCV.
- Lucas, Thomas, Konstantin Shmelkov, Karteek Alahari, Cordelia Schmid, and Jakob Verbeek (2019). "Adversarial training of partially invertible variational autoencoders". In: arXiv.
- Mathieu, Michael F, Junbo Jake Zhao, Junbo Zhao, Aditya Ramesh, Pablo Sprechmann, and Yann LeCun (2016). "Disentangling factors of variation in deep representation using adversarial training". In: NIPS.

References

- Netzer, Yuval, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng (2011). "Reading digits in natural images with unsupervised feature learning". In: NIPS-W.
- Olah, Chris, Alexander Mordvintsev, and Ludwig Schubert (2017). "Feature Visualization". In: Distill.
- Perarnau, Guim, Joost Van De Weijer, Bogdan Raducanu, and Jose M Álvarez (2016). "Invertible conditional gans for image editing". In: NIPS-W.
- Rasmus, Antti, Mathias Berglund, Mikko Honkala, Harri Valpola, and Tapani Raiko (2015). "Semi-supervised learning with ladder networks". In: NIPS.
- Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei (2015). "ImageNet Large Scale Visual Recognition Challenge". In: IJCV.
- Sajjadi, Mehdi, Mehran Javanmardi, and Tolga Tasdizen (2016). "Regularization With Stochastic Transformations and Perturbations for Deep Semi-Supervised Learning". In: NIPS.
- Shamir, O., S. Sabato, and Naftali Tishby (2010). "Learning and generalization with the information bottleneck". In: Theoretical Computer Science.
- Simonyan, Karen and Andrew Zisserman (2015). "Very deep convolutional networks for large-scale image recognition". In: ICLR.
- Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: JMLR.
- Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich (2015). "Going deeper with convolutions". In: CVPR.
- Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna (2016). "Rethinking the inception architecture for computer vision". In: CVPR.

References

- Tarvainen, Antti and Harri Valpola (2017). “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results”. In: NIPS.
- Tishby, Naftali, F. C. Pereira, and W. Bialek (1999). “The information bottleneck method”. In: Annual Allerton Conference on Communication, Control and Computing.
- Tishby, Naftali and Noga Zaslavsky (2015). “Deep learning and the information bottleneck principle”. In: Information Theory Workshop (ITW). IEEE.
- Tran, Luan, Xi Yin, and Xiaoming Liu (2017). “ Disentangled Representation Learning GAN for Pose-Invariant Face Recognition ”. In: CVPR.
- Wang, Ting-Chun, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro (2018). “High-resolution image synthesis and semantic manipulation with conditional gans”. In: CVPR.
- Wojna, Zbigniew, Vittorio Ferrari, Sergio Guadarrama, Nathan Silberman, Liang-Chieh Chen, Alireza Fathi, and Jasper Uijlings (2017). “The Devil is in the Decoder”. In: BMVC.
- Zhang, Chiyuan, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals (2017). “Understanding deep learning requires rethinking generalization”. In: ICLR.
- Zhang, Yuting, Kibok Lee, and Honglak Lee (2016). “Augmenting Supervised Neural Networks with Unsupervised Objectives for Large-scale Image Classification”. In: ICML.
- Zhao, Junbo, Michael Mathieu, Ross Goroshin, and Yann LeCun (2016). “Stacked What-Where Auto-encoders”. In: ICLR-W.

Appendix for SHADE

SHADE – Development details

- Layer-wise: $\Omega_{layers} = \sum_l \mathcal{H}(H_l|Y)$
- Unit-wise: $\Omega_{layers} < \Omega_{units} = \sum_l \sum_i \mathcal{H}\left(H_{l,i}\middle|Y\right)$
- Sufficient statistics assumption: $Y \rightarrow Z \rightarrow X \rightarrow H$
  $$\Rightarrow \mathcal{I}(H,Y) = \mathcal{I}(H,Z) \Rightarrow \mathcal{H}(H|Y) = \mathcal{H}(H|Z)$$
- Unit regularizer with Z:
  $\omega = \mathcal{H}(H|Y) = \mathcal{H}(H|Z) = \sum_z p(Z|H)\,\mathcal{H}(H|Z)$
- Variance bound: $\mathcal{H}(H|Z) < \frac{1}{2}\ln\left(2\pi e \mathcal{V}\mathrm{ar}(H|Z)\right)$
- Variance estimated using moving average of the expectation $\mathbb{E}(H|Z)$

SHADE – Implementation details

Algorithm A.1 Moving average updates: for $z \in \{0,1\}$, p^z estimates $p(Z=z)$ and μ^z estimates $\mathbb{E}(H \mid Z=z)$

Initialize: $\mu^0 = -1$, $\mu^1 = 1$, $p^0 = p^1 = 0.5$, $\lambda = 0.8$

for each mini-batch $\{h^{(k)}, k \in 1..K\}$ **do**

  for $z \in \{0,1\}$ **do**

    $p^z \leftarrow \lambda p^z + (1-\lambda)\frac{1}{K}\sum_{k=1}^{K} p(z \mid h^{(k)})$

    $\mu^z \leftarrow \lambda \mu^z + (1-\lambda)\frac{1}{K}\sum_{k=1}^{K} \frac{p(z \mid h^{(k)})}{p^z} h^{(k)}$

  end for

end for

$$\mathcal{V}\text{ar}(H \mid Z) = \int_{\mathcal{H}} p(h) \int_{\mathcal{Z}} p(z \mid h)\big(h - \mathbb{E}(H \mid z)\big)^2 \,\mathrm{d}z\,\mathrm{d}h$$

$$\approx \frac{1}{K}\sum_{k=1}^{K}\left[\int_{\mathcal{Z}} p(z \mid h^{(k)})\big(h^{(k)} - \mathbb{E}(H \mid z)\big)^2 \,\mathrm{d}\boldsymbol{z}\right] ;$$

$$\Omega_{\text{SHADE}} = \sum_{\ell=1}^{L}\sum_{i=1}^{D_\ell}\sum_{k=1}^{K}\sum_{z\in\{0,1\}} p\left(Z_{\ell,i} = z \,\middle|\, h_{\ell,i}^{(k)}\right)\left(h_{\ell,i}^{(k)} - \mu_{\ell,i}^{z}\right)^2 .$$

SHADE – ImageNet results

SCIENCES SORBONNE UNIVERSITÉ

	Accuracy (%)	
	Top-1	Top-5
ResNet-101	77.56%	93.89%
WELDON	78.51%	94.65%
WELDON + SHADE	**80.14%**	**95.35%**

SHADE – Binary model hypothesis

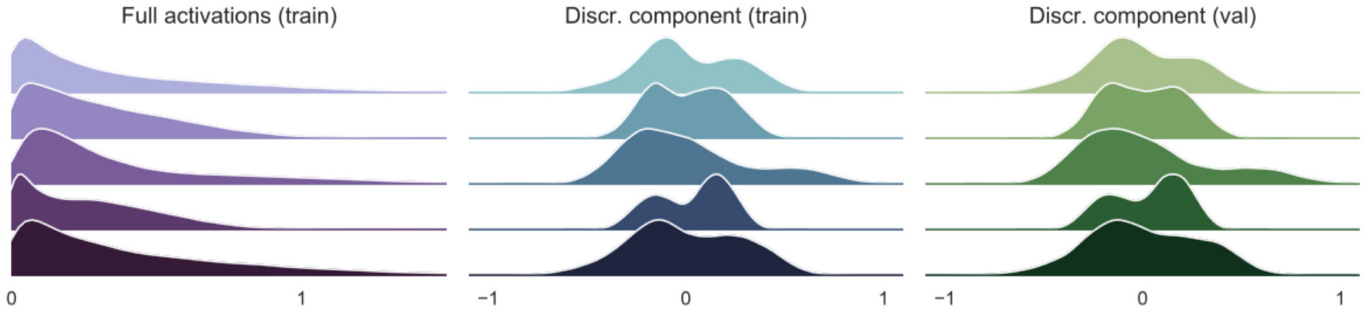


Figure A.2. – **Visualization of 5 neurons from the penultimate activations** (*i.e.* the input of the last fully-connected layer) of an Inception model trained on CIFAR-10. On the left is the distribution of the values taken by each neuron H. In the middle and right is the distribution of the discriminiative component H^* of the neuron (the part that does not belong to the kernel of the layer weights).

	Original	Binarized layer		
Architecture	score	Before $\hat{y}$ ($\mathbf{h}_{L-1}$)	Middle ($\mathbf{h}_{L/2}$)	After input ($\mathbf{h}_1$)
MLP	64.68	64.92	62.45	61.13
AlexNet	83.25	82.71	82.38	82.01
Inception	91.34	91.41	90.88	90.21
ResNet	93.24	92.67	92.09	91.99

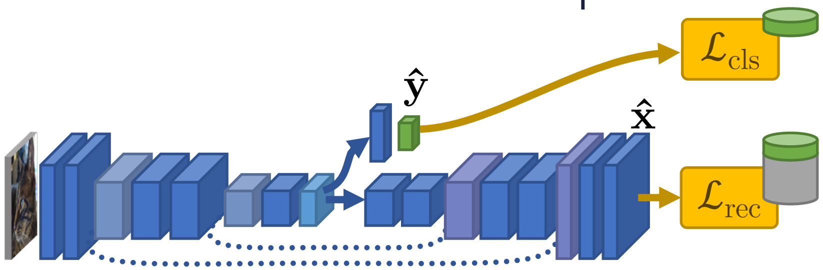
Table A.1. – **Classification accuracy (%) using binarized activation** on CIFAR-10 test set.

Appendix for HybridNet

Related work

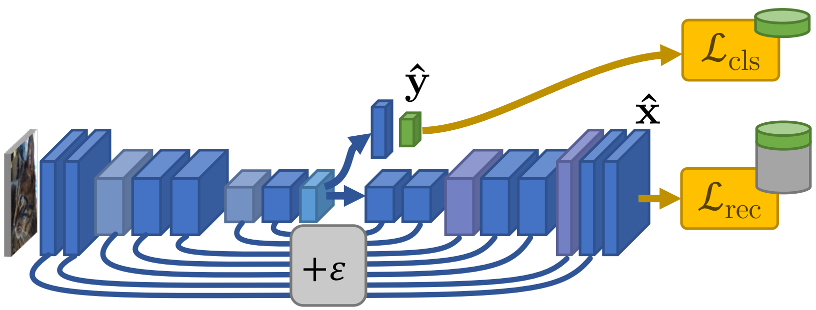
Reconstruction-based
⇒ information skip

SWWAE (Zhao, 2016)



Pooling location info

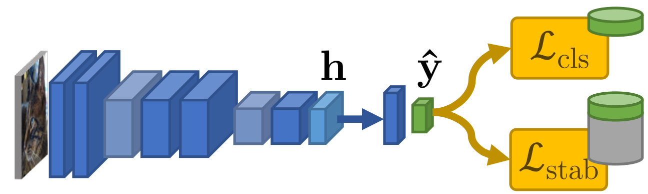
LadderNet (Rasmus, 2015)



Noisy representations info.

Stability-based

Mean Teacher (Tarvainen, 2017)



→ Enforces invariance to sources of random variability

→ Create virtual targets $\mathbf{z}^{(i)}$, e.g. avg of outputs

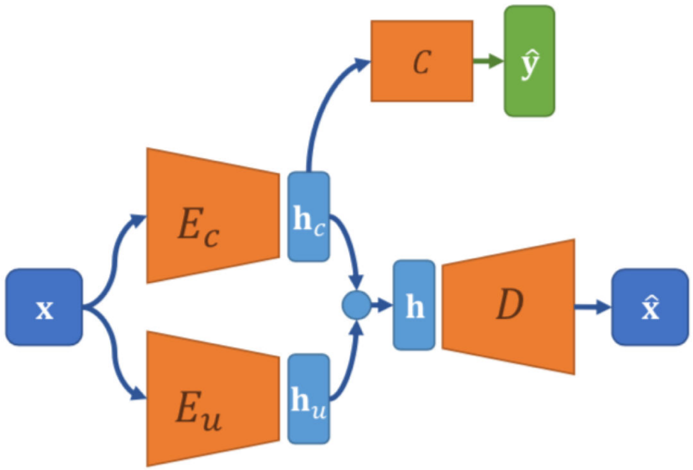
$$\mathcal{L}_{\text{stab}} = \left\|\hat{\mathbf{y}}^{(i)} - \mathbf{z}^{(i)}\right\|_2$$

Limits

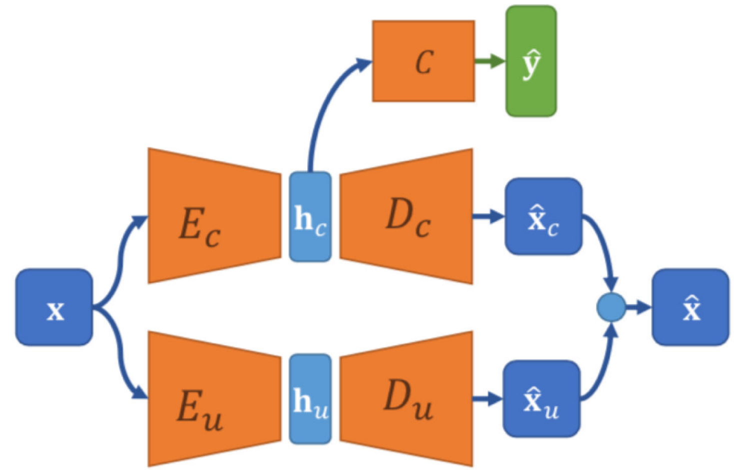
Fixed type of skipped information

Does not encourage extraction of more generic patterns

HybridNet – Fusion strategies



(a) **Early fusion** merging the representations and using a single decoder.

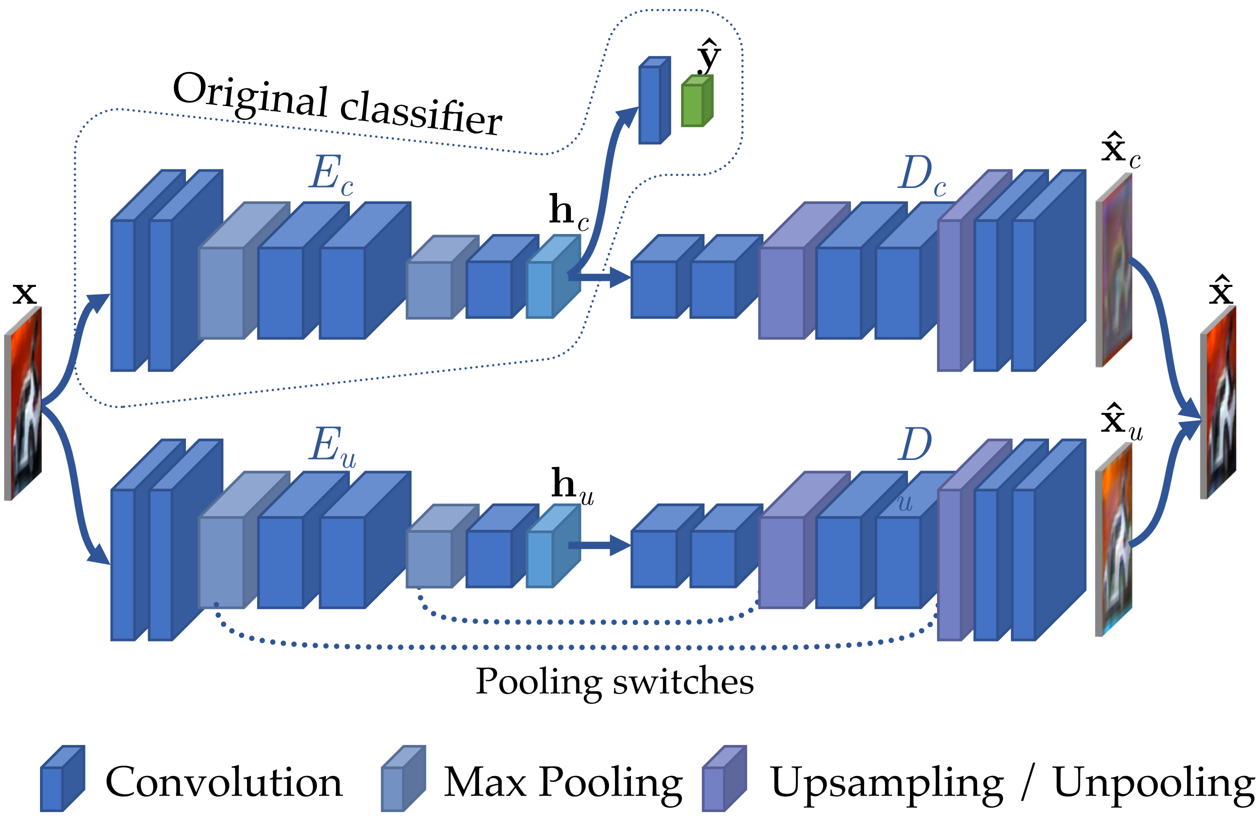


(b) **Late fusion** using two decoders and merging the reconstructions.

- → Richer interactions
- → More complex to control

- → More simple interactions
- → Possible to control each branch directly

HybridNet – Architecture example



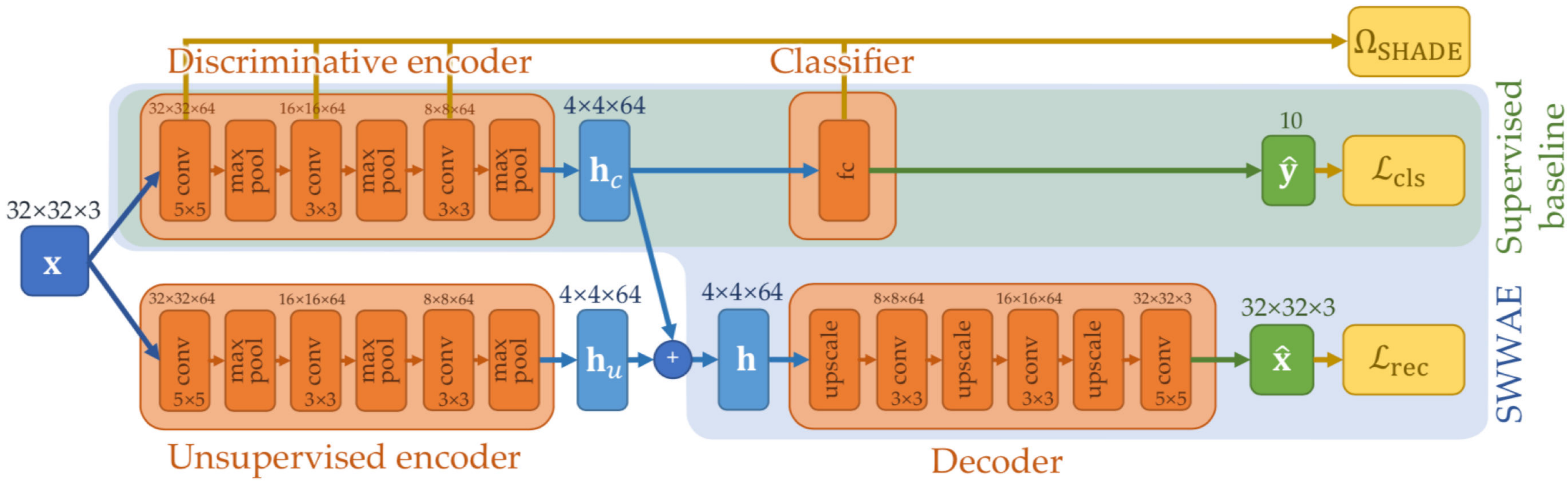
HybridNet – Ablation study

SCIENCES SORBONNE UNIVERSITÉ

Model	Labeled samples N_s		
	1000	2000	4000
Classification	63.4	71.5	79.0
Classification and stability	65.6	74.6	81.3
Auto-encoder	65.0	73.6	79.8
Auto-encoder and stability	71.8	80.4	84.9
HybridNet architecture	63.2	74.0	80.3
HybridNet architecture and full training loss	**74.1**	**81.6**	**86.6**

CIFAR-10, ConvLarge

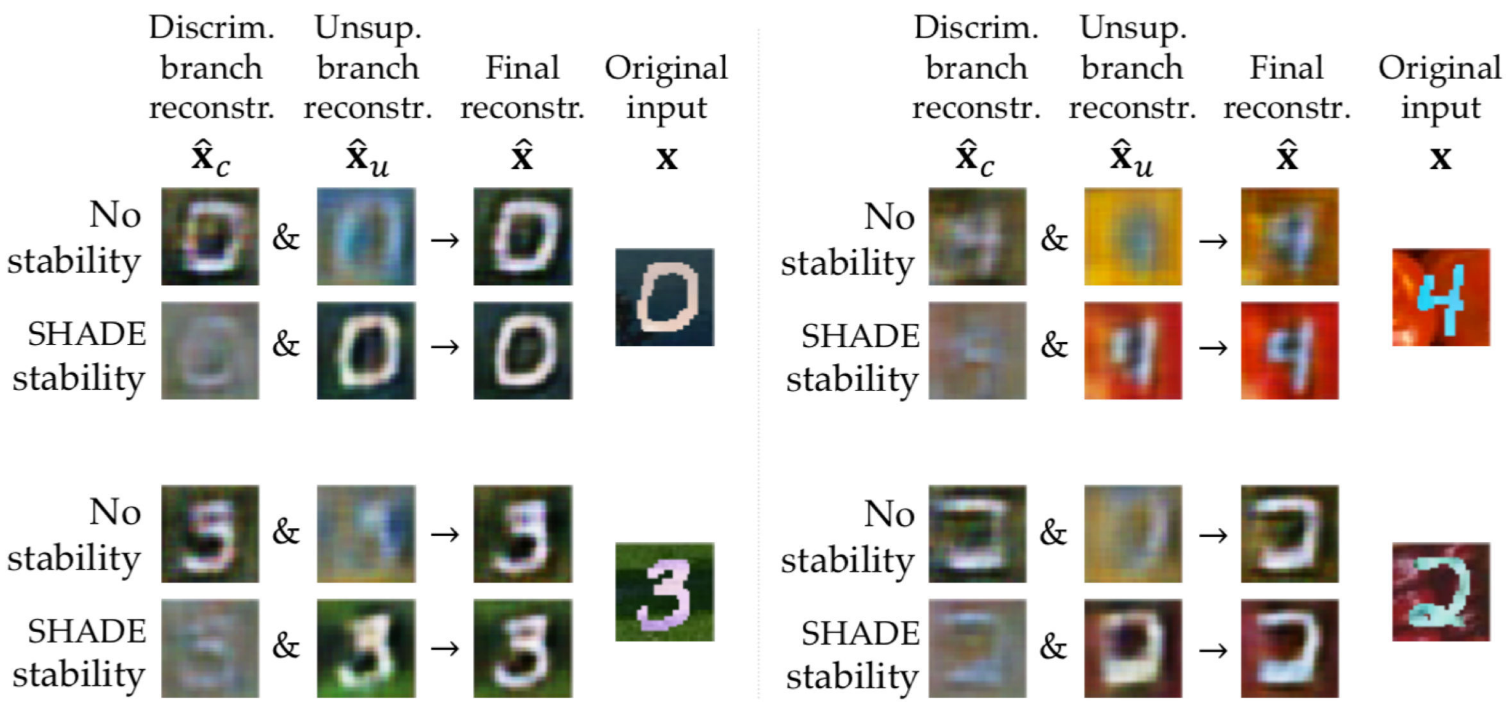
HybridNet – Architecture with SHADE



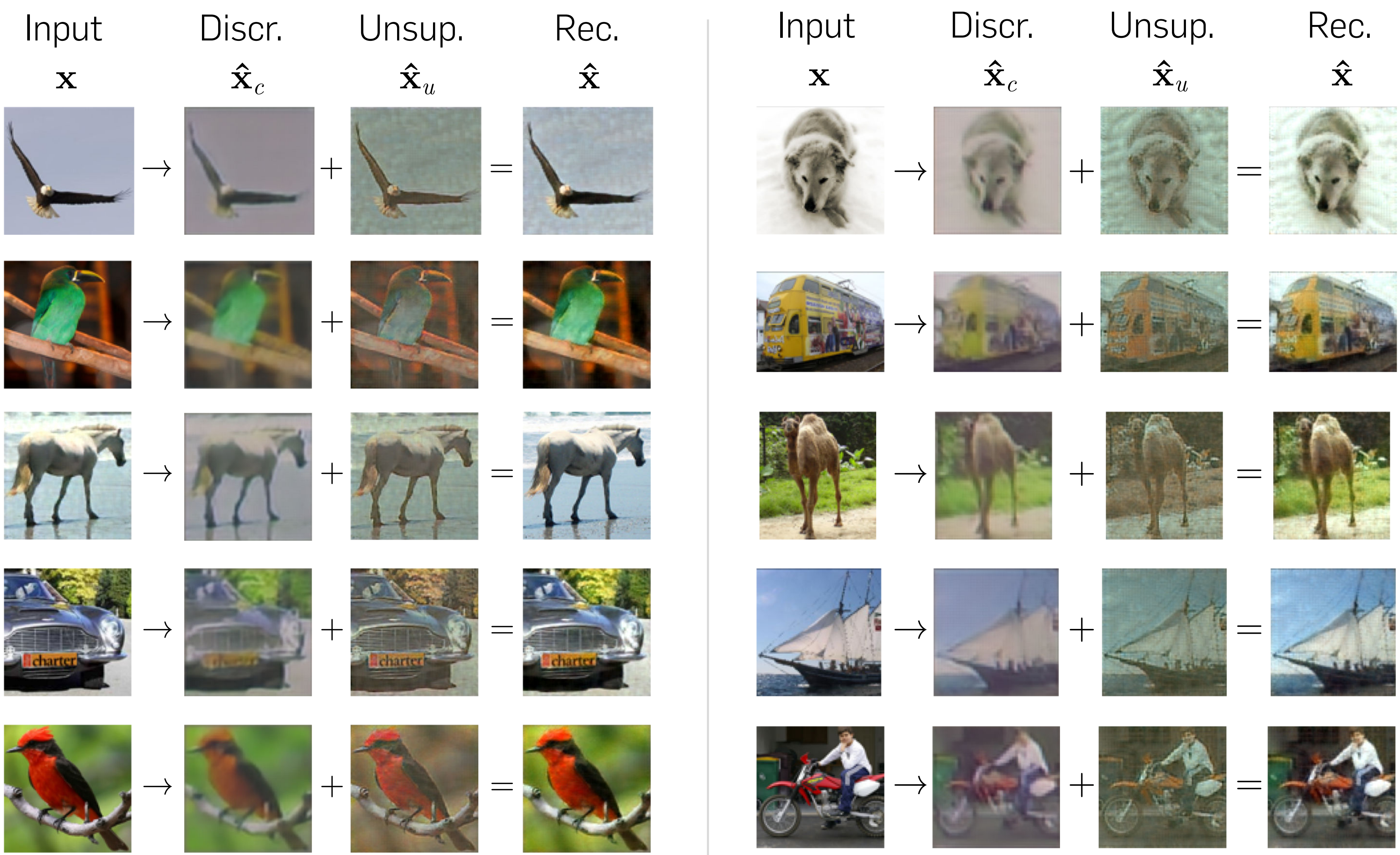
HybridNet – SHADE results

SCIENCES SORBONNE UNIVERSITÉ

Dataset	MNIST		MNIST-M		SVHN
Nb labeled samples N_s	100	1000	100	1000	1000
Supervised baseline	83.26	95.51	47.14	83.09	75.03
SWWAE*	86.38	95.72	45.83	82.89	75.27
HybridNet no regul.	84.13	96.01	48.07	84.86	75.63
HybridNet + weight decay	87.71	95.98	48.62	83.69	76.13
HybridNet + SHADE	**89.15**	**97.18**	**52.58**	**88.23**	**79.12**



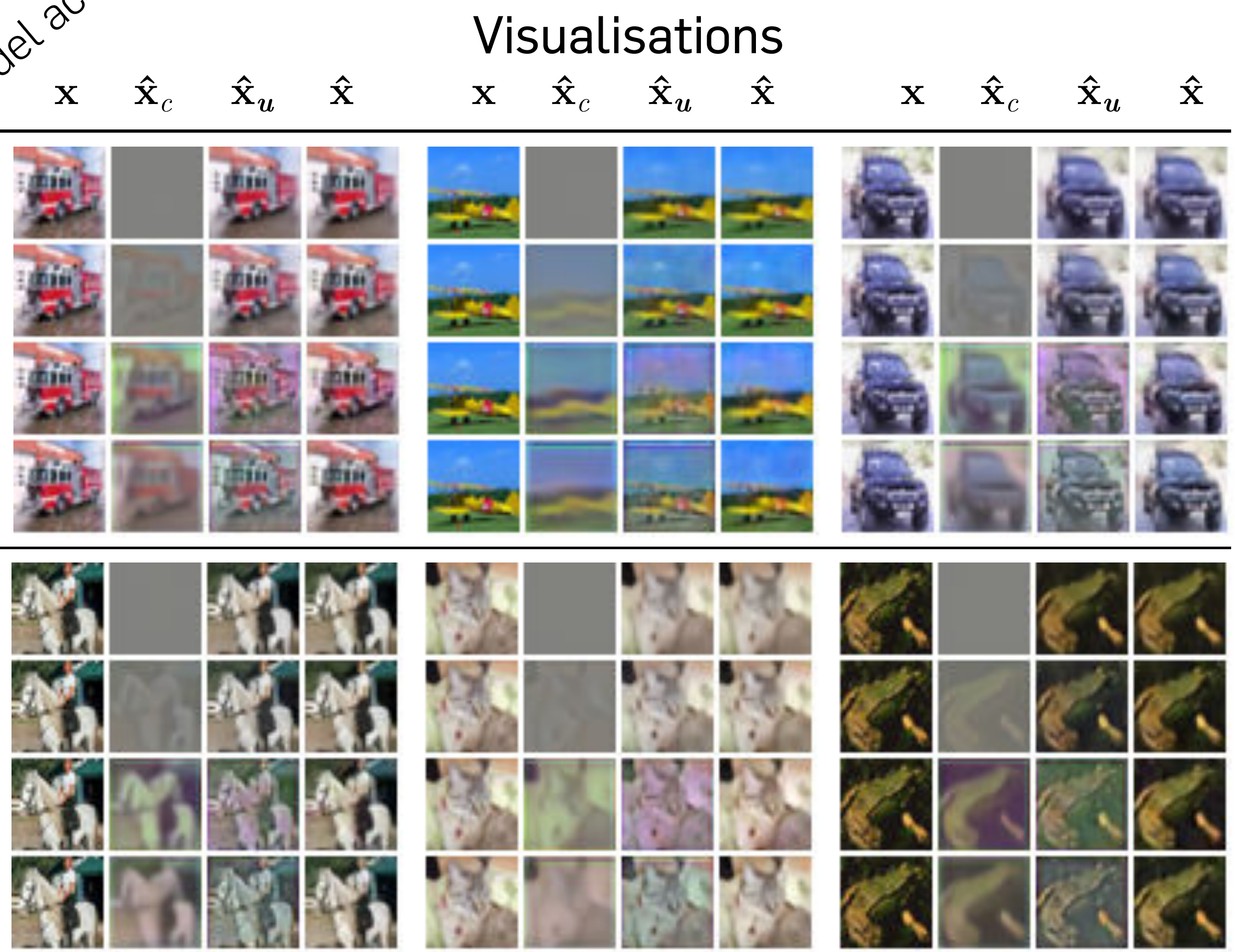
HybridNet – Visual results



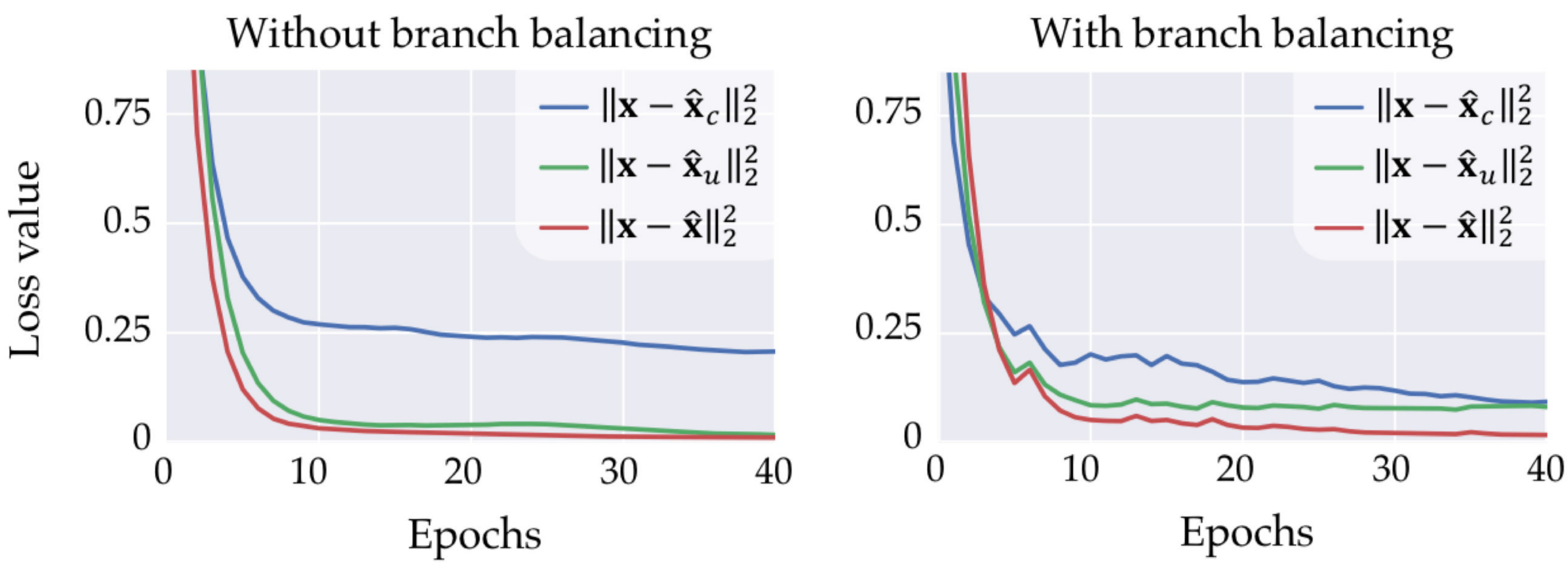
HybridNet – Ablation study, visual analysis

SCIENCES SORBONNE UNIVERSITÉ

Final rec.	Intermed. rec.	Complement.	Stability	Model accuracy
✓				72.4
✓	✓			74.0
✓	✓	✓		75.2
✓	✓	✓	✓	81.6
✓				72.4
✓	✓			74.0
✓	✓	✓		75.2
✓	✓	✓	✓	81.6



HybridNet – Branch balancing effect on loss



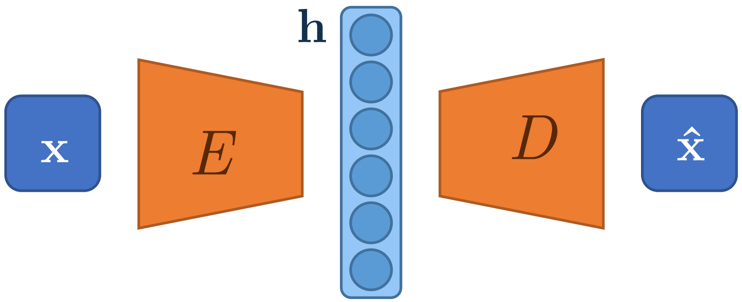
Appendix for DualDis

DualDis – Related work

SCIENCES SORBONNE UNIVERSITÉ

Unsupervised disentangling

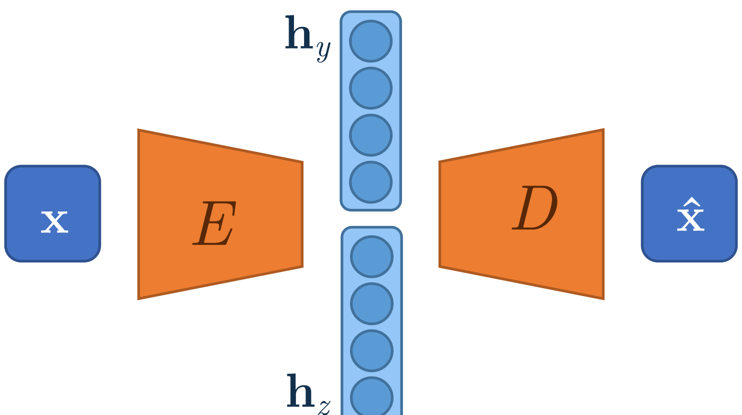
(Higgins, 2017; Chen, 2018; Kim, 2018; Dupont, 2018; Hu, 2018; ...)



- → Produce independent neurons or groups of neurons
- → No interpretation of the neurons' role
- → Specific metrics (~ verify 1 neuron for each labeled factor)

Supervised disentangling

(Perarnau, 2016; Lample, 2017; Mathieu, 2016; Klys, 2018; Hadad, 2018; Liu, 2018; ...)



- → Separate 2 information domains
- → Use labels for 1 or 2 domains
- → Used for discriminative and generative tasks

DualDis – GAN & VAE

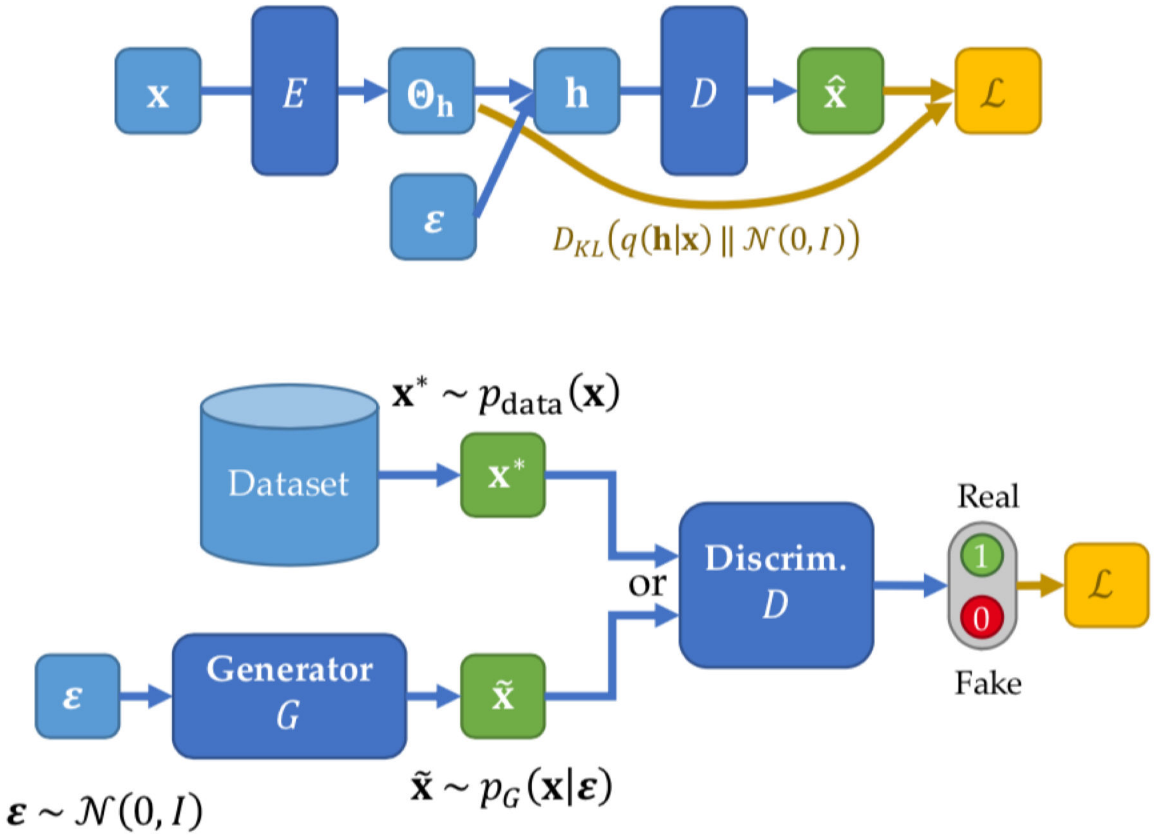


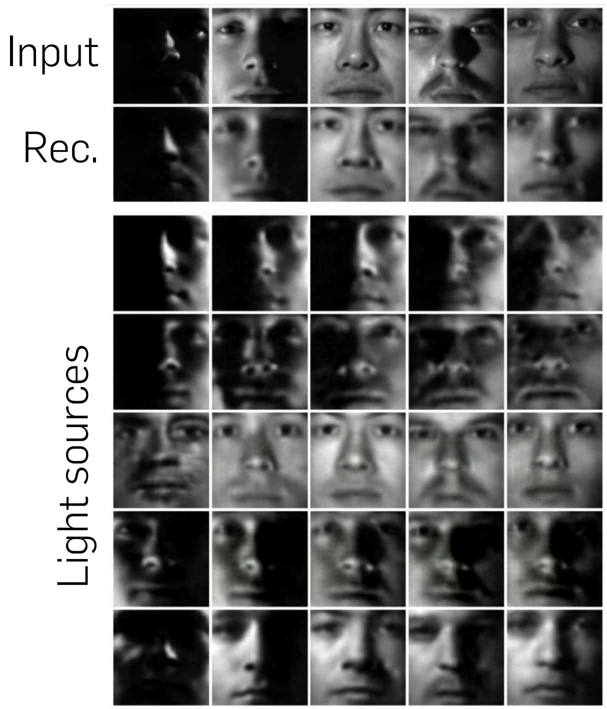
Figure 4.2. – Schematic representation of a VAE (top) and a GAN (bottom).

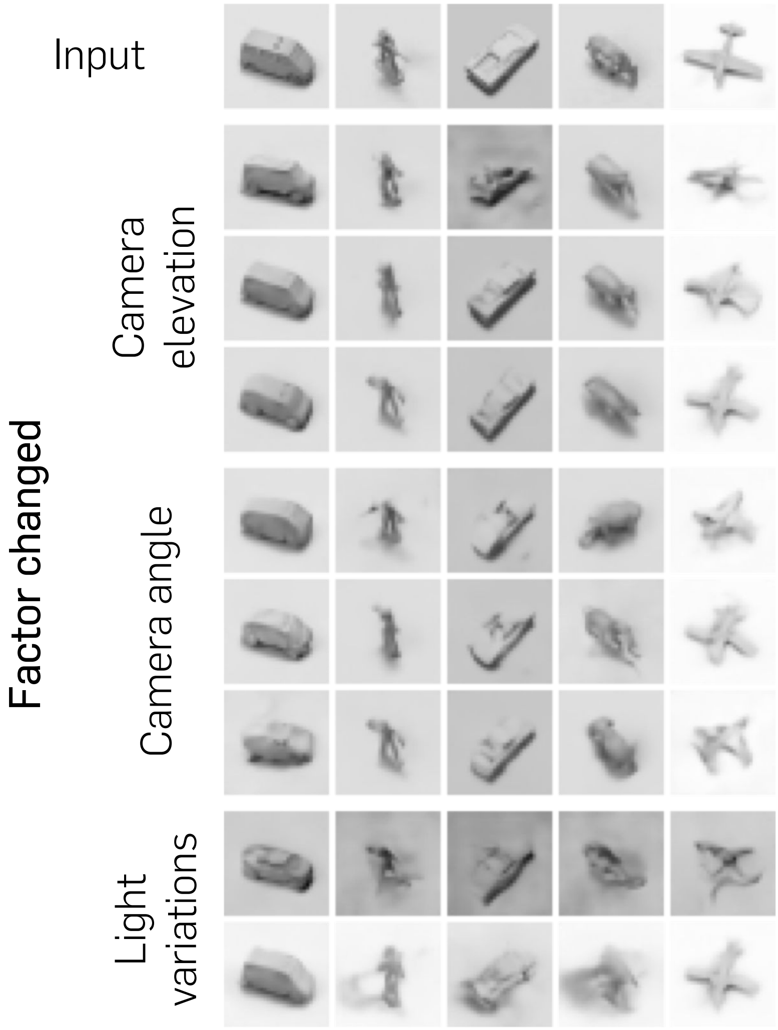
DualDis – Quantitative results

		Model	Labels used	Aggr. Metric	Accuracy $\mathbf{h}_y \rightarrow \mathbf{y}$	Accuracy $\mathbf{h}_z \rightarrow \mathbf{z}$	Disentangling $\mathbf{h}_z \rightarrow \boldsymbol{y}_{\text{adv}}$	Disentangling $\mathbf{h}_y \rightarrow \mathbf{z}_{\text{adv}}$
CelebA	(A)	Multi-task classif.	$\mathbf{y}, \mathbf{z}$	61.1	**77.6%**	**91.8%**	65.5%	9.5%
	(B)	HybridNet-like	$\mathbf{y}$	65.1	73.0%	82.4%	95.5%	9.4%
	(B')	HybridNet-like + attr	$\mathbf{y}, \mathbf{z}$	65.2	72.7%	90.1%	88.5%	9.5%
	(C)	MTAN	$\mathbf{y}, \mathbf{z}, \mathbf{z}_{\text{test}}$	—	68.9%	—	—	13.8%
	(D)	UAI adv. loss	$\mathbf{y}$	63.7	67.9%	80.3%	**97.3%**	9.3%
	(D')	UAI adv. loss + attr	$\mathbf{y}, \mathbf{z}$	65.0	68.0%	89.4%	92.9%	9.5%
	(E)	Adv. on y only	$\mathbf{y}$	64.7	69.2%	83.6%	96.4%	9.6%
		DualDis	$\mathbf{y}, \mathbf{z}$	**68.0**	71.1%	88.6%	**97.3%**	**14.9%**
Yale-B	(A)	Multi-task classif.	$\mathbf{y}, \mathbf{z}$	81.5	98.5%	**97.2%**	85.3%	45.1%
	(B)	HybridNet-like	$\mathbf{y}$	65.3	97.6%	93.7%	23.3%	46.5%
	(B')	HybridNet-like + attr	$\mathbf{y}, \mathbf{z}$	80.5	**99.0%**	96.9%	80.0%	46.1%
	(C)	MTAN	$\mathbf{y}, \mathbf{z}, \mathbf{z}_{\text{test}}$	—	98.4%	—	—	70.3%
	(D)	UAI adv. loss	$\mathbf{y}$	60.0	98.6%	65.5%	28.1%	48.0%
	(D')	UAI adv. loss + attr	$\mathbf{y}, \mathbf{z}$	65.1	96.1%	95.8%	44.4%	24.1%
	(E)	Adv. on y only	$\mathbf{y}$	79.8	98.3%	84.1%	92.5%	44.4%
		DualDis	$\mathbf{y}, \mathbf{z}$	**92.0**	98.6%	**97.3%**	**98.8%**	**73.4%**
NORB	(A)	Multi-task classif.	$\mathbf{y}, \mathbf{z}$	53.7	93.0%	84.2%	13.5%	24.0%
	(B)	HybridNet-like	$\mathbf{y}$	51.1	93.3%	76.8%	12.2%	22.1%
	(B')	HybridNet-like + attr	$\mathbf{y}, \mathbf{z}$	52.5	92.9%	84.1%	10.7%	22.2%
	(C)	MTAN	$\mathbf{y}, \mathbf{z}, \mathbf{z}_{\text{test}}$	—	92.2%	—	—	**30.5%**
	(D)	UAI adv. loss	$\mathbf{y}$	51.8	92.8%	76.0%	13.7%	24.7%
	(D')	UAI adv. loss + attr	$\mathbf{y}, \mathbf{z}$	52.5	93.2%	82.8%	8.0%	26.0%
	(E)	Adv. on y only	$\mathbf{y}$	67.3	92.2%	76.9%	78.9%	21.1%
		DualDis	$\mathbf{y}, \mathbf{z}$	**72.3**	**93.5%**	**84.5%**	**80.7%**	**30.5%**

DualDis – Image editing

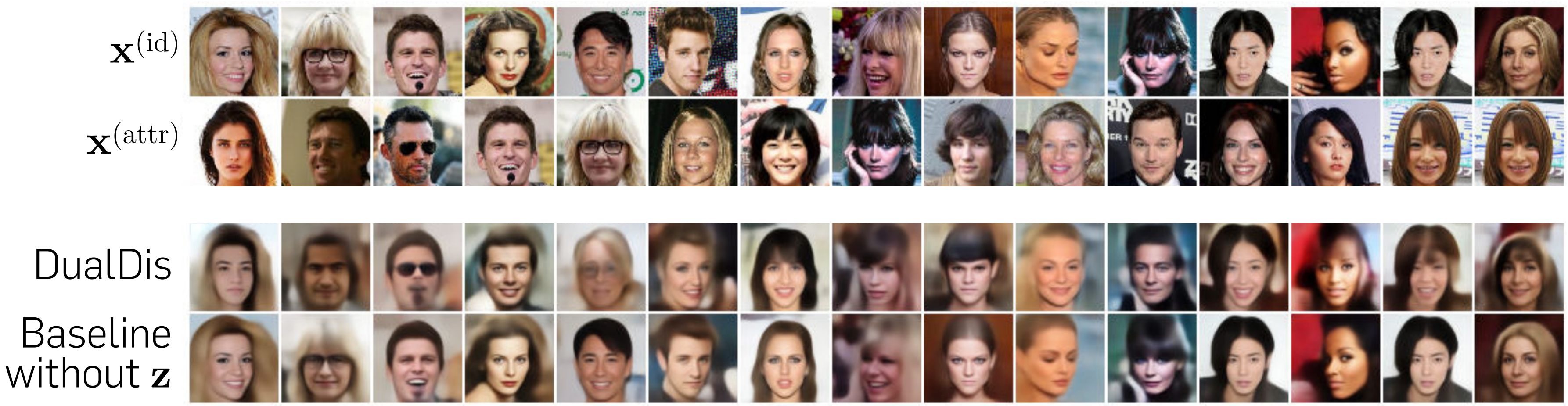
DualDis – Visual results on Yale and NORB





DualDis – Disentangling visualization

Initial images. $\mathbf{x}^{(\mathrm{id})}$: Identity source / $\mathbf{x}^{(\mathrm{attr})}$: Attribute source



Generations from $\left(\mathbf{h}_y^{(\mathrm{id})}, \mathbf{h}_z^{(\mathrm{attr})}\right)$ produced by DualDis and the baseline