Medical Image Analysis with Deep Learning
-
Uncertainties - Explainability

M. Sdika [1]

[1] Univ.Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS

UMR 5220, U1206, F-69100, LYON, France

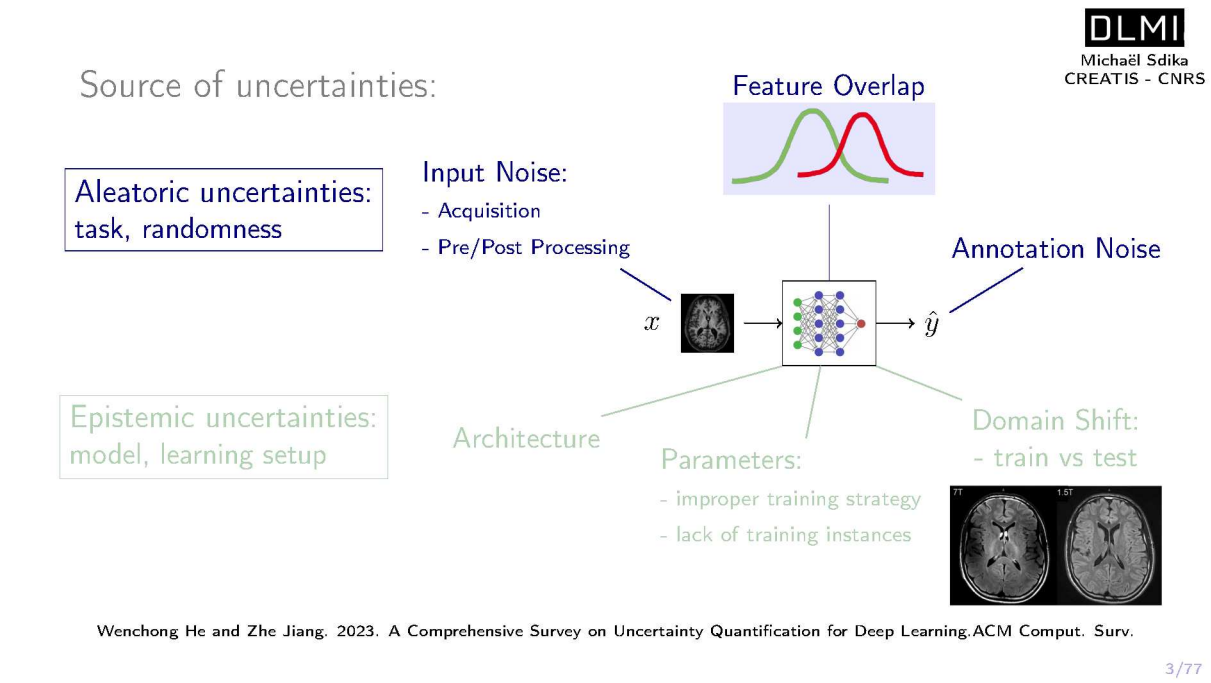
Deep learning

- Deep network:
  - high performance
  - model developped from the data
    $\rightarrow$ black box, lack of transparency
- decision on human people health $\rightarrow$ need safeguard
  - interpretation, explaination, localization
  - uncertainties estimation

Source of uncertainties:

Aleatoric uncertainties:
task, randomness

Epistemic uncertainties:
model, learning setup

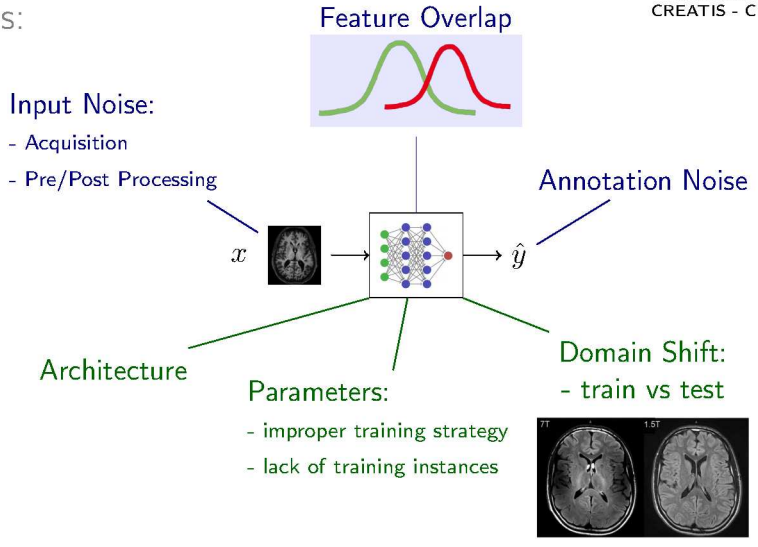


Wenchong He and Zhe Jiang. 2023. A Comprehensive Survey on Uncertainty Quantification for Deep Learning.ACM Comput. Surv.

Source of uncertainties:

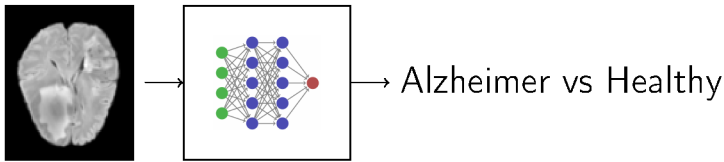
Aleatoric uncertainties:
task, randomness

Epistemic uncertainties:
model, learning setup



Wenchong He and Zhe Jiang. 2023. A Comprehensive Survey on Uncertainty Quantification for Deep Learning.ACM Comput. Surv.

Domain Shift



Covariate shift:

- network input
- change in acquisition
- demographics

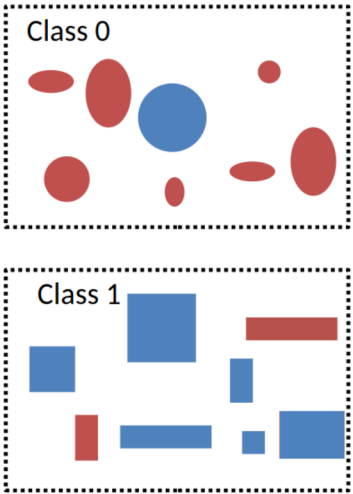
Label shift
Concept Drift:

- labels repartition

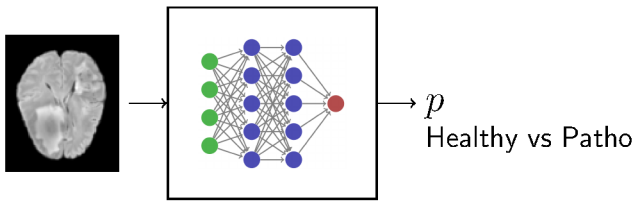
Semantic shift:

- network output
- input HIV+ patient
- OOD

Dataset biases, Confounder



Dataset biases, Confounder



Do the "Healthy" and "Patho" data match in:

- imaging parameters: scanner brand ? acquisition parameters ? ...
- population: age ? morphology ? ...
- something else ...

Content

DLMI
Michaël Sdika
CREATIS - CNRS

Content

DLMI
Michaël Sdika
CREATIS - CNRS

Content

DLMI
Michaël Sdika
CREATIS - CNRS

Scoring Rule

Scoring rule: $S(p, y)$
measures how well an estimated probability vector p explains the observed labels y

Strictly proper scoring rule:
- proper: minimal for the true distribution of y
- strictly proper: the true distribution of y is the only minimum

Example:
- cross entropy (CE), negative log likelyhood (NLL)

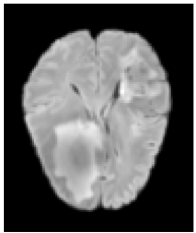
$$-\log(p_y)$$

- Brier Score:

$$\frac{1}{C}\sum_c (p_c - y_c)^2$$

DLMI
Michaël Sdika
CREATIS - CNRS

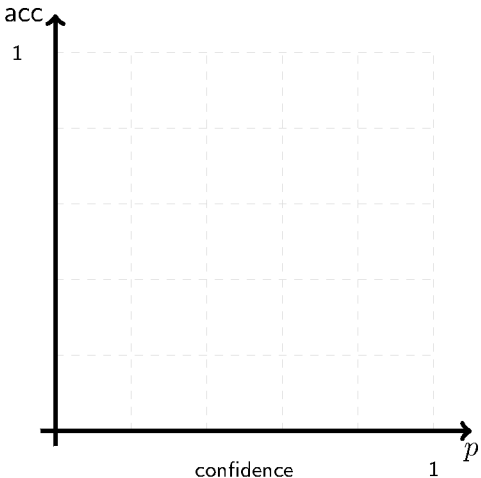
The network ouput as an uncertainty measure ??



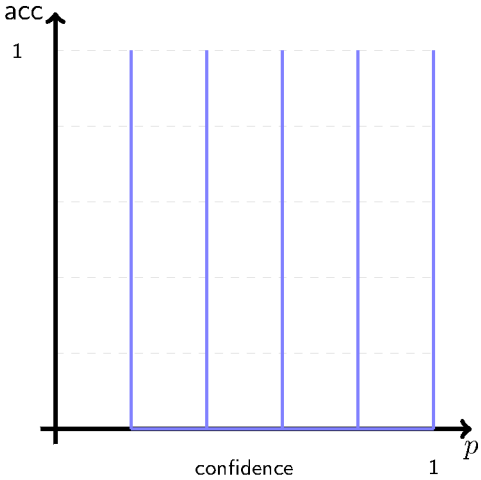
Is it the probability to give a correct output ?

$\longrightarrow$ **network confidence**

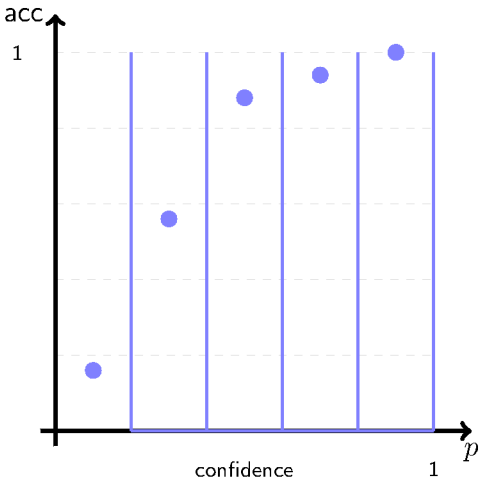
Reliability Diagram



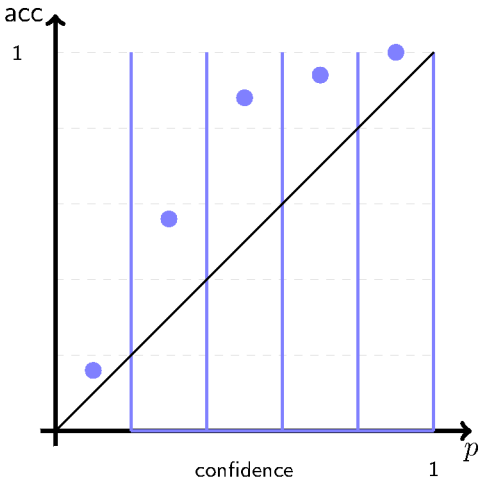
Reliability Diagram



Reliability Diagram

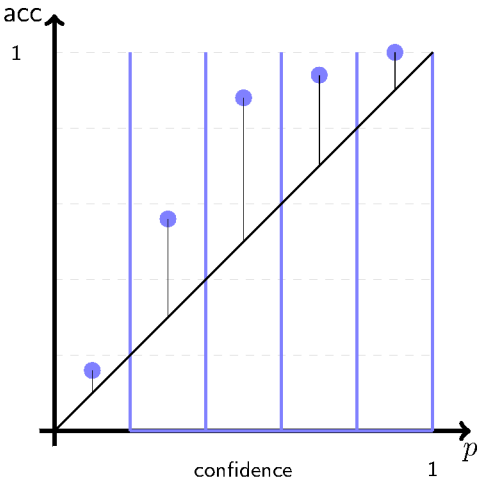


Reliability Diagram



DLMI
Michaël Sdika
CREATIS - CNRS

Reliability Diagram



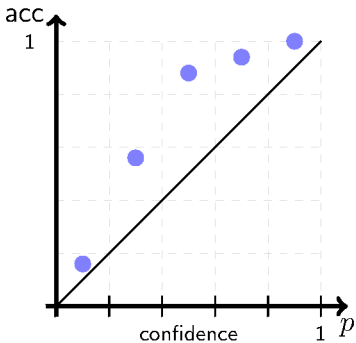
Expected Calibration Error (ECE)

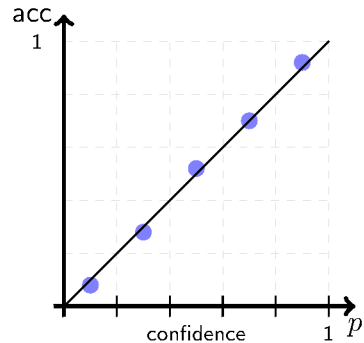
$$\text{ECE} = \frac{1}{B}\sum_{b=1}^{B} |\text{conf}_b - \text{acc}_b|$$

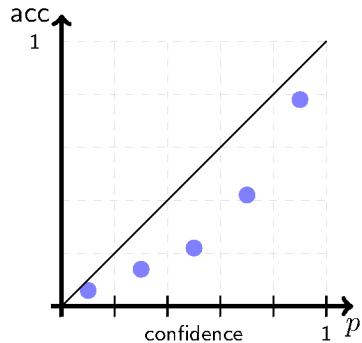
Multi-class, either:

- avg accuracy VS avg confidence
- classwize reliability diagram

Reliability Diagram

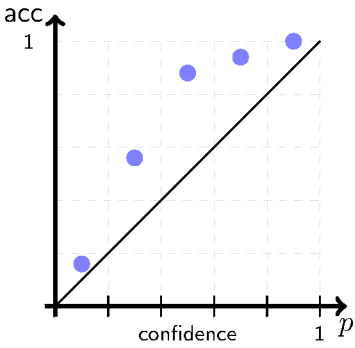


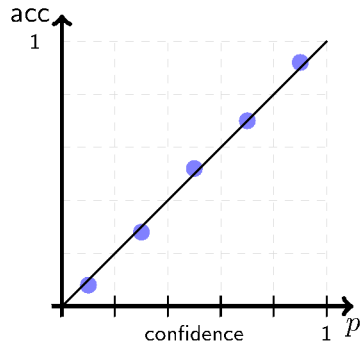


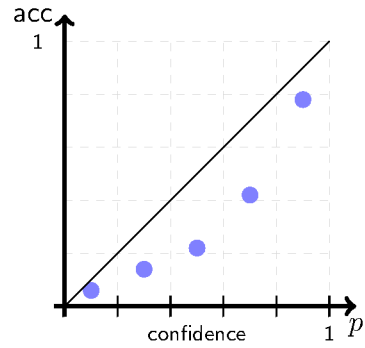


DLMI
Michaël Sdika
CREATIS - CNRS

Reliability Diagram



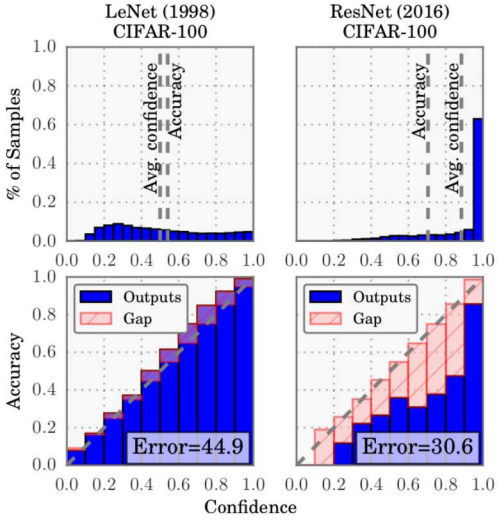


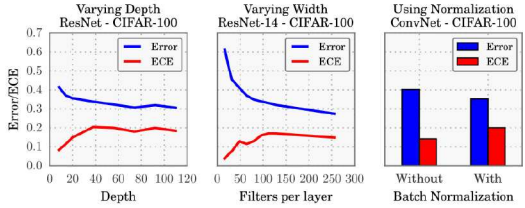


under-confident **calibrated** **over-confident**

DLMI
Michaël Sdika
CREATIS - CNRS

Modern deep networks are not calibrated





cause:

- depth/width/batch norm
- Cross Entropy loss:
  - $-\log(p_y)$

Guo *et. al.* 2017, On Calibration of Modern Neural Networks

DLMI
Michaël Sdika
CREATIS - CNRS

Label Smoothing

y GT label

correct with probability $1 - \epsilon$

$$CE = -\sum_i y_i \log(p_i)$$

- standard : $y_i = 0/1$
- label smoothing: $y_i = \frac{\epsilon}{K-1}$ / $1 - \epsilon$

Szegedy et al, CVPR 2016, Rethinking the inception architecture for computer vision

Deep Learning, Goodfellow et al

DLMI
Michaël Sdika
CREATIS - CNRS

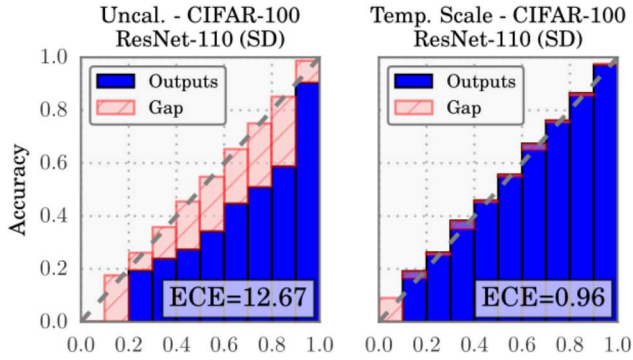
Temperature scaling

Softmax :

$$\frac{e^{z_k}}{\sum e^{z_i}} \longrightarrow \frac{e^{z_k/T}}{\sum e^{z_i/T}}$$

On the validation dataset:

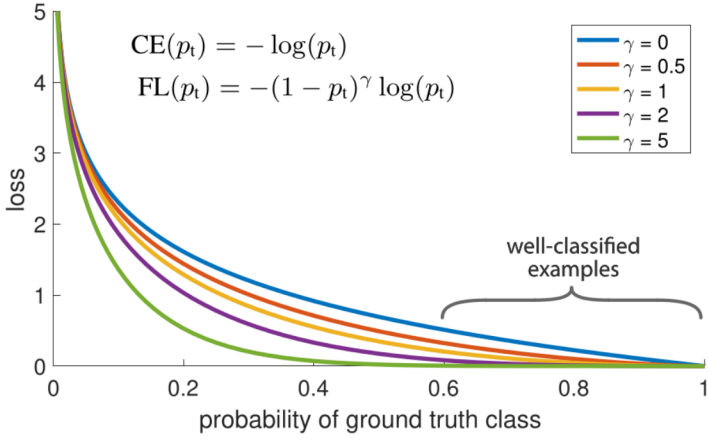
$$\min_T \mathrm{NLL}(T)$$



Guo *et. al.* 2017, On Calibration of Modern Neural Networks

DLMI
Michaël Sdika
CREATIS - CNRS

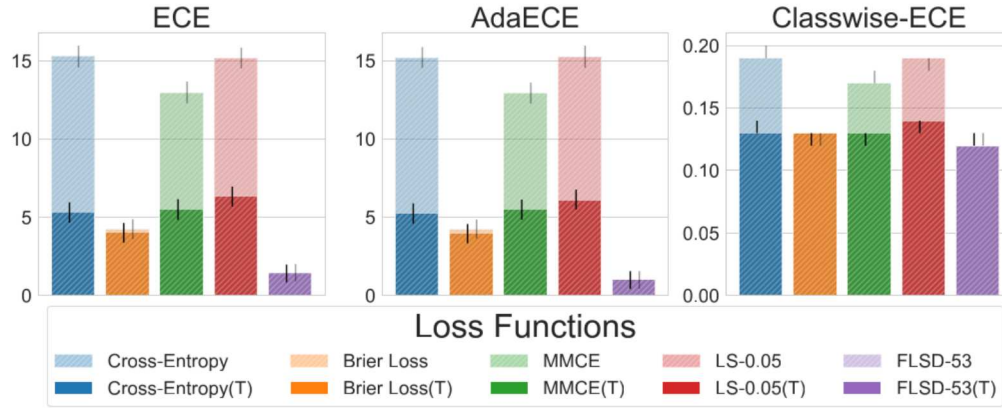
Focal Loss



Lin etal, ICCV 2017. Focal loss for dense object detection
Mukhoti etal Neurips 2020. Calibrating deep neural networks using focal loss

DLMI
Michaël Sdika
CREATIS - CNRS

Focal Loss



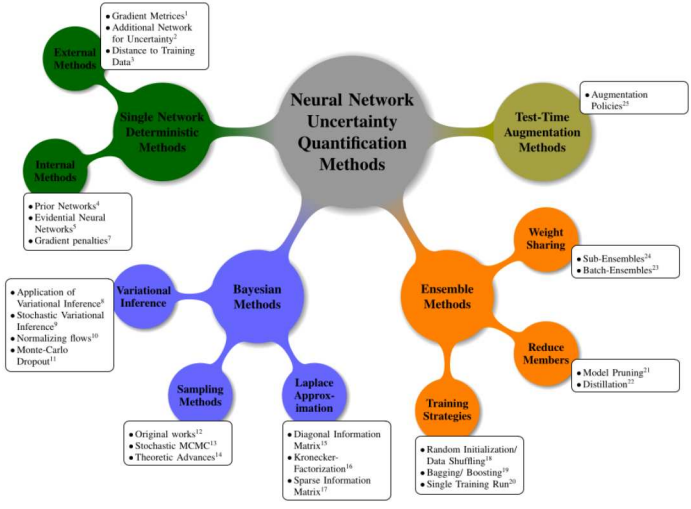
Lin etal, ICCV 2017. Focal loss for dense object detection
Mukhoti etal Neurips 2020. Calibrating deep neural networks using focal loss

DLMI
Michaël Sdika
CREATIS - CNRS

Content

DLMI
Michaël Sdika
CREATIS - CNRS

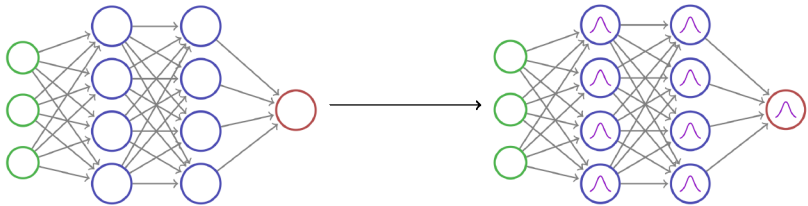
Uncertainties approaches



Gawlikowski etal, Artificial Intelligence Review 2023, A survey of uncertainty in deep neural networks

DLMI
Michaël Sdika
CREATIS - CNRS

Bayesian Networks



Inference: MC sampling

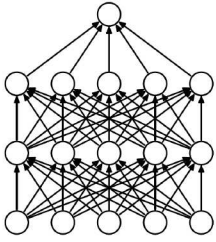
- D: training dataset
- draw random $w_k \sim q(w)$
- $p(y|x, D) \approx \frac{1}{K} \sum_{k=1}^{K} p(y|x, w_k)$

Uncertainty Estimation:

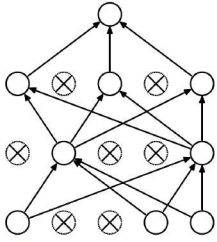
- $p^c = p(y = c|x, w_k)$
- $H[y|x, D] \approx -\sum_c p^c \log(p^c)$

DLMI
Michaël Sdika
CREATIS - CNRS

Dropout Regularization

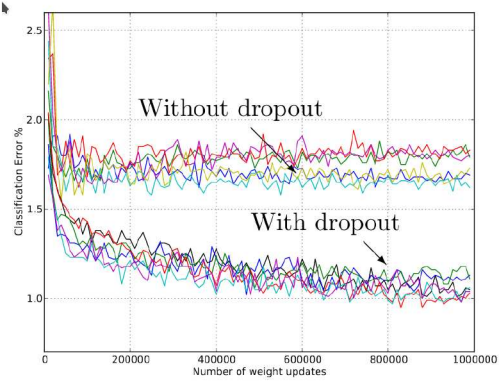


(a) Standard Neural Net



(b) After applying dropout.

Random set units to 0 during training



Test error with/without dropout

Srivastava *et. al.* 2014, Dropout: A Simple Way to Prevent Neural Networks from Overfitting

DLMI
Michaël Sdika
CREATIS - CNRS

MC Dropout for uncertainty estimation

network: $x \rightarrow y = y(x)$

- dropout during inference: sample several outputs $(y_i(x))_{1,N}$
- Gal 2016: these samples are $\approx$ drawn from $p(\hat{y}|x^*)$
- estimate first and second moments of $p(y|x)$

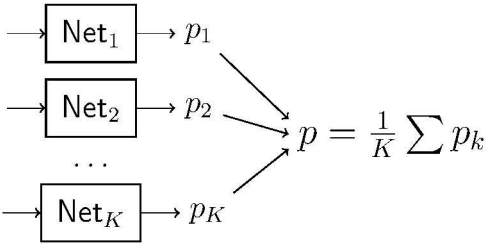
$$E(y|x) \approx SampleMean\,(y_i)$$

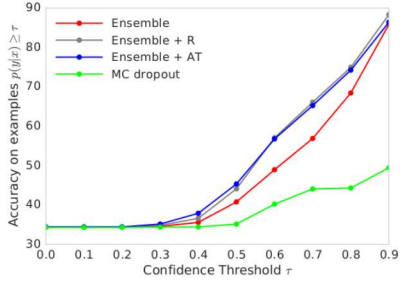
$$Var(y|x) \approx \tau I + \mathrm{SampleVar}(y_i)$$

Gal & Ghahramani, ICML 2016: Dropout as a bayesian approximation: representing model uncertainty in deep learning

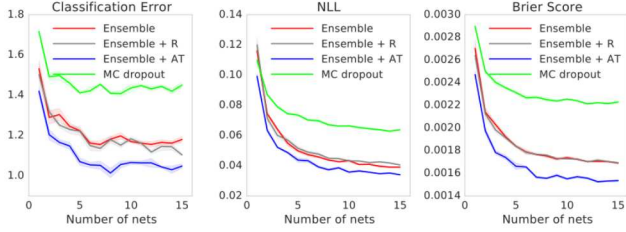
DLMI
Michaël Sdika
CREATIS - CNRS

Deep ensemble





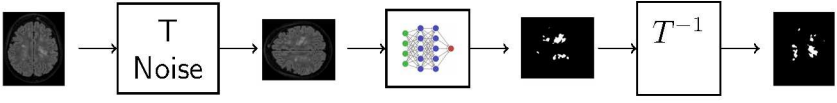
- train a set of network
- stochasticity from random init
- in practice, K=5



Lakshminarayanan, *et. al.* NIPS 2017, Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles

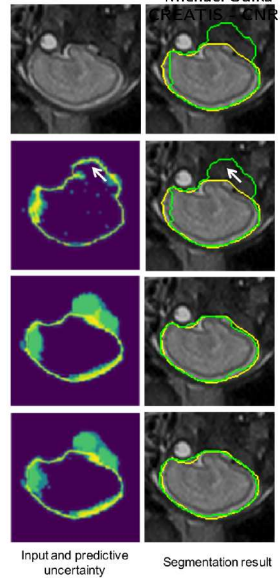
DLMI
Michaël Sdika
CREATIS - CNRS

Test Time Augmentation (TTA)



During test:
- generate multiple output
- prediction: mean / mode
- uncertainty: std-dev / entropy

aleatoric uncertainty related to
image transformations and noise



Wang et al, 2019, Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with CNN.

DLMI
Michaël Sdika
CREATIS - CNRS

Content

DLMI
Michaël Sdika
CREATIS - CNRS

Conformal Prediction

{ fox squirrel 0.99 }

{ fox squirrel 0.82, gray fox, 0.03, bucket, 0.02, rain barrel 0.02 }

{ marmot 0.30, fox squirrel 0.22, mink 0.18, weasel 0.16, beaver 0.03, polecat 0.01 }

for a test image I, find a set of labels $C(I)$ s.t.:

$$P\left(y \in C(I)\right) \geq 1 - \alpha$$

Angelopoulos & Bates, A gentle introduction to conformal prediction and distribution-free uncertainty quantification." arXiv 2021

DLMI
Michaël Sdika
CREATIS - CNRS

Conformal Prediction: Solving the problem

- heuristic uncertainty measure for the model ex: $\rightarrow \hat{f}(x)$
- Define a score function $s(x, y)$ (lower=better) ex: $\rightarrow s(x, y) = 1 - \hat{f}_y(x)$
- Define the set C_q:

$$C_q(x) = \{y \text{ s.t } s(x, y) < q\}$$

- Notice that

$$P\left(y \in C_q(x)\right) = P\left(s(x, y) < q\right) = F_S(q)$$

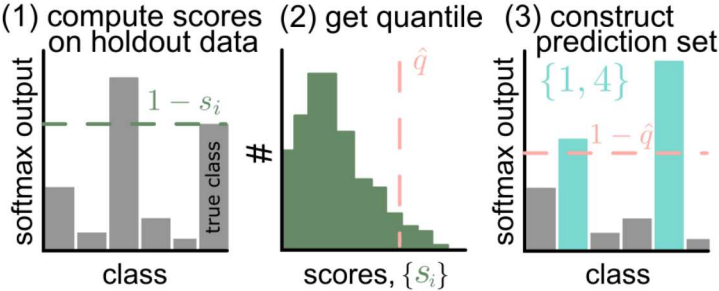
is the cumulative distribution function of $s(X, Y)$

$$\Rightarrow \hat{q} \approx (1 - \alpha)\text{-quantile}$$

Angelopoulos & Bates, A gentle introduction to conformal prediction and distribution-free uncertainty quantification." arXiv 2021

DLMI
Michaël Sdika
CREATIS - CNRS

Conformal Prediction: Solving the problem



Calibration data:

- conformal score:

$$s_i = 1 - f_{y_i}(x_i)$$

- $q \approx (1 - \alpha)$-quantile of scores

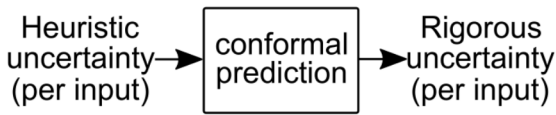
$x \in$ TEST (y unknown):

$$C(x) = \left\{ y | 1 - \hat{f}_y(x) \leq q \right\}$$

Angelopoulos & Bates, A gentle introduction to conformal prediction and distribution-free uncertainty quantification." arXiv 2021

DLMI
Michaël Sdika
CREATIS - CNRS

Conformal Prediction



- Limits
  - bad model / bad heuristic uncertainties / small calib set
    $\rightarrow$ return all the labels
  - sensitive to domain shift between calib / test

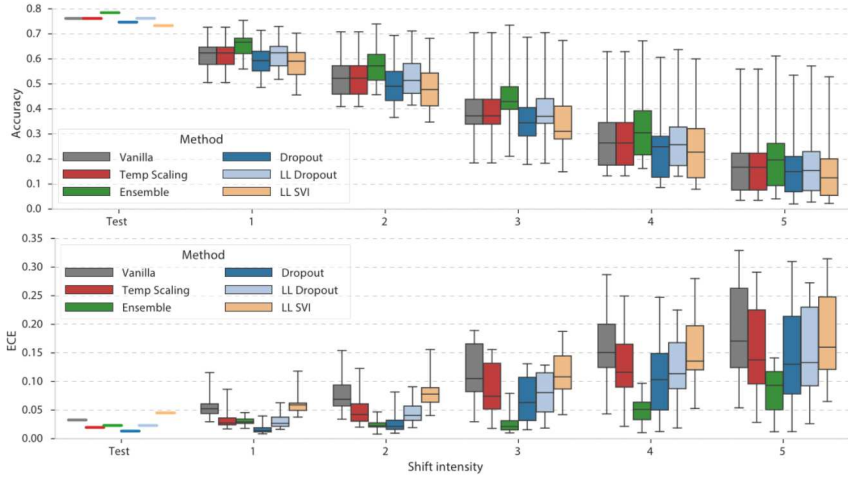
Angelopoulos & Bates, A gentle introduction to conformal prediction and distribution-free uncertainty quantification." arXiv 2021

DLMI
Michaël Sdika
CREATIS - CNRS

Content

DLMI
Michaël Sdika
CREATIS - CNRS

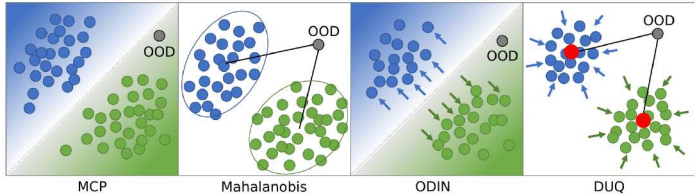
Uncertainties estimation is not robust to domain shift



Ovadia, et al. 2019 "Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift."

DLMI
Michaël Sdika
CREATIS - CNRS

Some confidence and distance based OOD



Maximum Class Probability

ID : $\max_c p_c > \text{thresh}$

MCP with better calibration:

- MCP Dropout
- MCP Deep ensemble

ODIN

- $\tilde{x}$: adv attack to increase confidence
- adjust temperature (G-ODIN:large)
- score : MCP ($\tilde{x}$)

DUQ

- feature learning
- contrastive learning
- score : RBF kernel distance to centroid

KNN

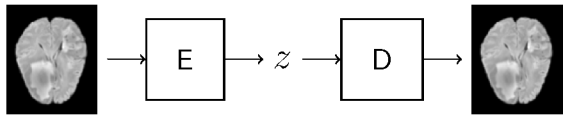
- contrastive
- KNN

Mahalanobis

Berger, et al. UNSURE MICCAI 2021, Confidence-based out-of-distribution detection: a comparative study and analysis
Hendrycks & Gimpel, ICLR 2017, A baseline for detecting misclassified and out-of-distribution examples in neural networks
Liang etal ICLR 2018: Enhancing the reliability of out-of-distribution image detection in neural networks
Van Amersfoort etal, ICML 2020 : Uncertainty estimation using a single deep deterministic neural network
Sun, et al, ICML 2022, Out-of-distribution detection with deep nearest neighbors

DLMI
Michaël Sdika
CREATIS - CNRS

Reconstruction-based OOD



Assumption: better reco for ID than OOD

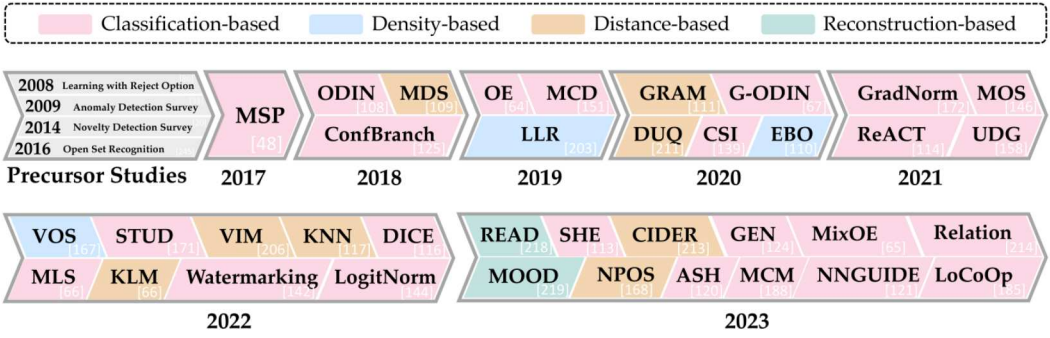
score: $\|\hat{x} - x\|$

$$\|\hat{x} - x\| + \lambda(z - \mu)\Sigma^{-1}(z - \mu)$$

Denouden etal, arxiv 2018, Improving Reconstruction Autoencoder Out-of-distribution Detection with Mahalanobis Distance

DLMI
Michaël Sdika
CREATIS - CNRS

ODD

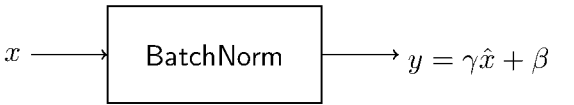


Yang et al., arXiv, Generalized out-of-distribution detection: A survey

Zhang et al, arxiv, OpenOOD v1.5: Enhanced Benchmark for Out-of-Distribution Detection

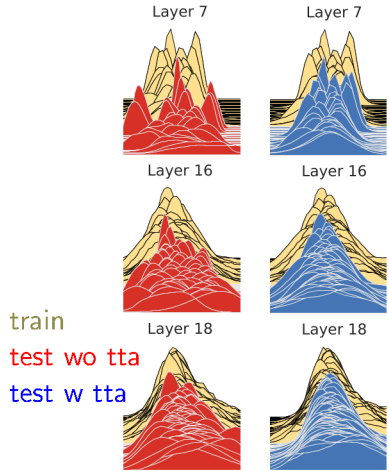
DLMI
Michaël Sdika
CREATIS - CNRS

Test Time Adaptation with BatchNorm



TTA flavors:

- adjust μ/σ on each batch at test
- adjust on full test domain
- fine tune γ/β
  with self sup. / consistency loss



Nado etal, ICML 2020, Evaluating prediction-time batch normalization for robustness under covariate shift
Liang etal 2024, A Comprehensive Survey on Test-Time Adaptation under Distribution Shifts
Wu etal 2024, Test-Time Domain Adaptation by Learning Domain-Aware Batch Normalization

DLMI
Michaël Sdika
CREATIS - CNRS

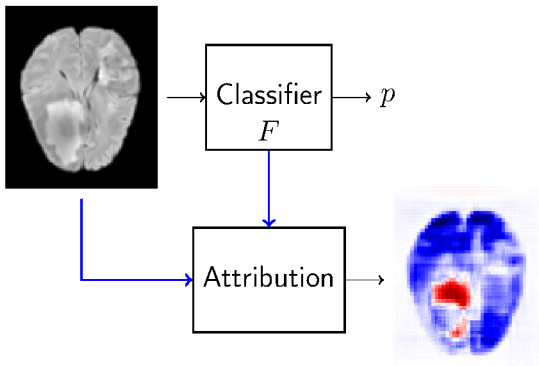
Content

DLMI
Michaël Sdika
CREATIS - CNRS

Content

DLMI
Michaël Sdika
CREATIS - CNRS

Attribution maps: WHERE, not WHAT

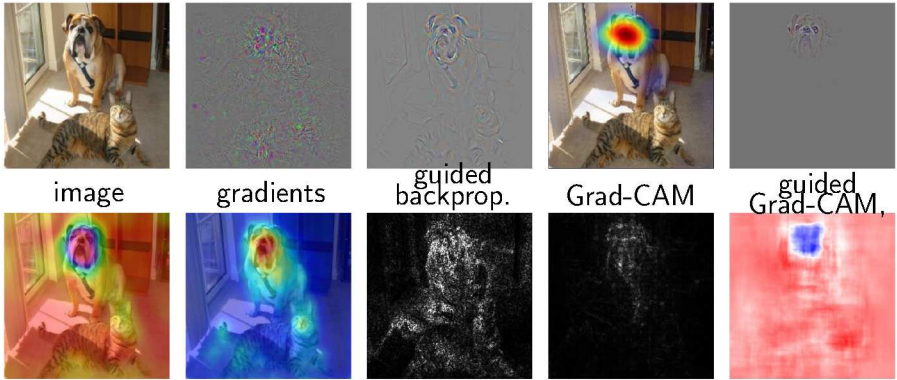


Post hoc: apply it once the network has been trained

blue: push toward negative
red : push toward positive

DLMI
Michaël Sdika
CREATIS - CNRS

Lots of methods, several approaches

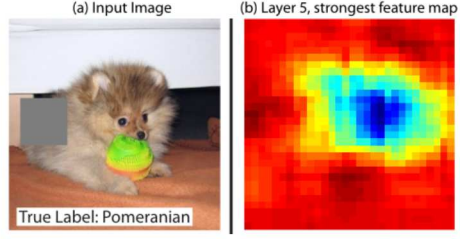


Perturbation // Gradient // Activation

Kumar & Kamath. "Attribution-based XAI methods in computer vision: A review." arXiv:2211.14736 (2022).

DLMI
Michaël Sdika
CREATIS - CNRS

Occlusion



$$A(I, B)(x) = F(I) - F(I^x)$$

I^x: baseline value at pixel x
otherwize I

- I^x might be OOD
- untractable $\Rightarrow$ patch, superpixel...

Zeiler & Fergus, ECCV 2014, Visualizing and Understanding Convolutional Networks

DLMI
Michaël Sdika
CREATIS - CNRS

Saliency maps: the gradient as an attribution maps

Saliency Maps / Gradient:

$$A(I) = \nabla_I F$$

Input x Gradient:

$$A(I) = (I - B)\nabla_I F$$

$y = w_1 f_1 + w_2 f_1 + w_3 f_3$

G: w_1 (local)
IxG: $w_1 f_1$ (global)

saliency map

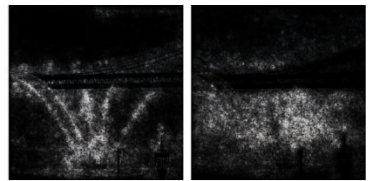
Simonyan et al, Deep inside convolutional networks: Visualising image classification models and saliency maps. ICLR 2014
Shrikumar et al, Learning important features through propagating activation differences. ICML 2017

DLMI
Michaël Sdika
CREATIS - CNRS

Path Methods: Integrated Gradient

$$A(I,B) = (I-B)\int_0^1 \frac{\partial F}{\partial I}\left(B+\alpha(I-B)\right)d\alpha$$

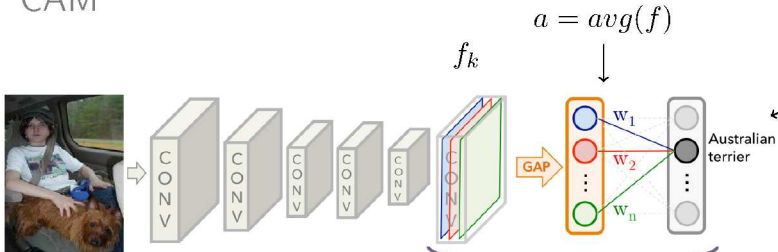
satisfy: Linearity w.r. model, Sensitivity, Implementation invariance
Completeness ($\sum_x A(I)_x = F(I) - F(B)$)



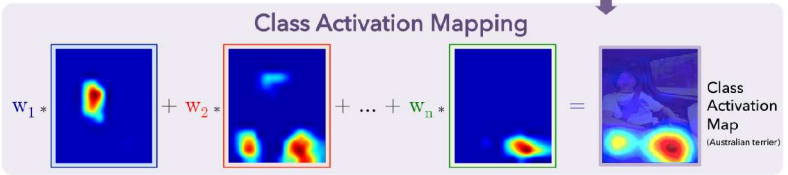
M Sundararajan, et al ICML 2017, Axiomatic attribution for deep networks

DLMI
Michaël Sdika
CREATIS - CNRS

CAM



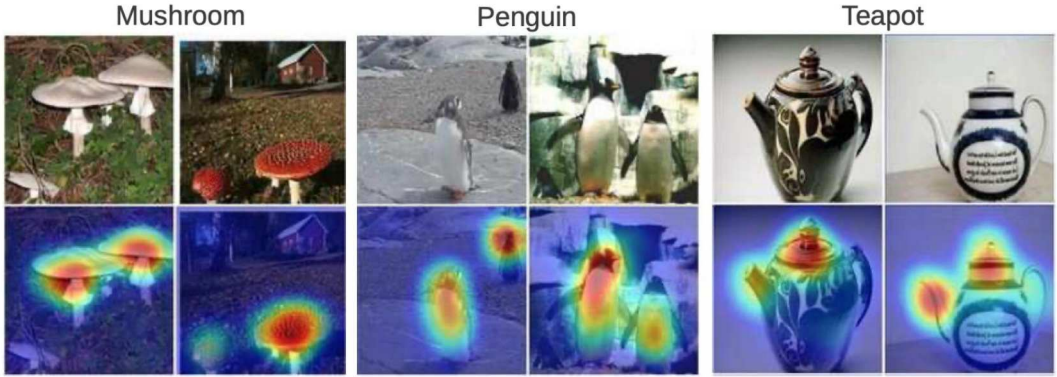
$$y = \sum_i w_i a_i$$
$$\propto \sum_{i,x,y} w_i f_i(x,y)$$
$$\propto \sum_{x,y} \sum_i w_i f_i(x,y)$$
$$\propto \sum_{x,y} A(x,y)$$



Zhou et al, CVPR 2016, Learning deep features for discriminative localization

DLMI
Michaël Sdika
CREATIS - CNRS

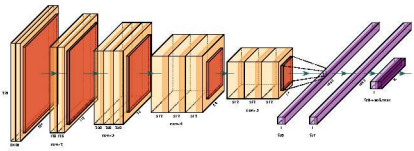
CAM



Zhou et al, CVPR 2016, Learning deep features for discriminative localization

DLMI
Michaël Sdika
CREATIS - CNRS

Grad-CAM

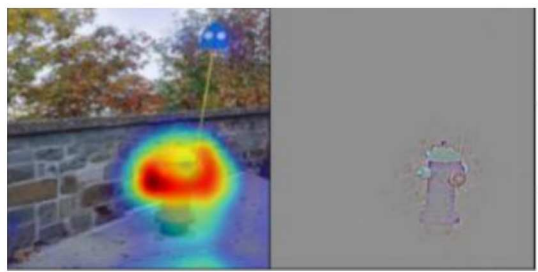


Can be used on any layer

(usually last conv)

$$w_k = \text{avg}_X \left(\frac{\partial y}{\partial f_k} \right)$$

$$A(I) = \text{relu} \left(\sum_k w_k f_k(X) \right)$$

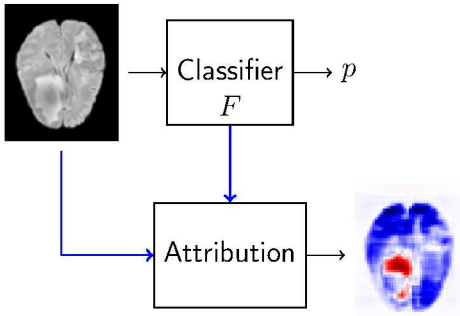


GradCAM Guided GradCAM

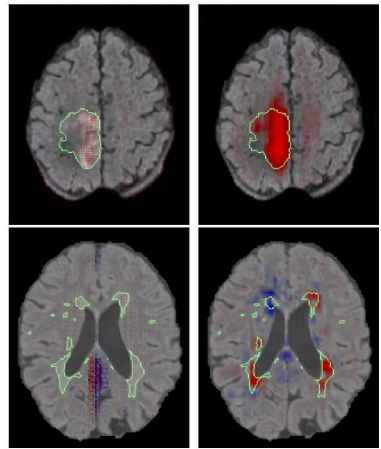
Selvaraju et al, ICCV 2017, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization

DLMI
Michaël Sdika
CREATIS - CNRS

Attribution maps constrained training



Constraint:
I is Healthy $\Rightarrow A(I)$ is blue



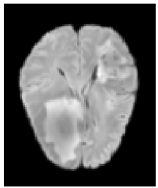
naive constrained

DLMI
Michaël Sdika
CREATIS - CNRS

Content

DLMI
Michaël Sdika
CREATIS - CNRS

Counterfactual Explanation



Minimal changes in the input data to switch class ?

Naive:

$$\min_{F(x')=1-y} \|x' - x\|$$

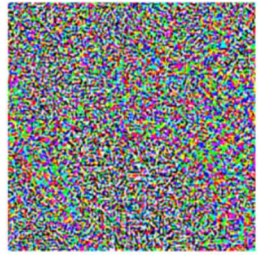
DLMI
Michaël Sdika
CREATIS - CNRS

Naive = Adversarial attack

$\boldsymbol{x}$

"panda"
57.7% confidence

$+ .007 \times$



$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

$=$

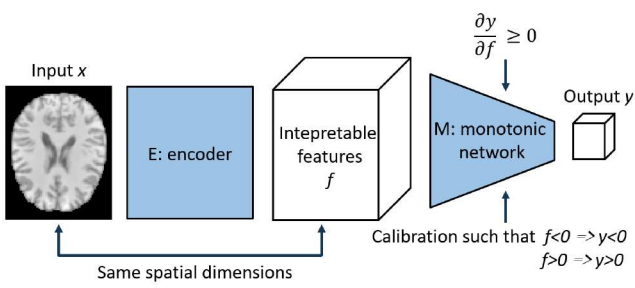
$\boldsymbol{x} + \epsilon \text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

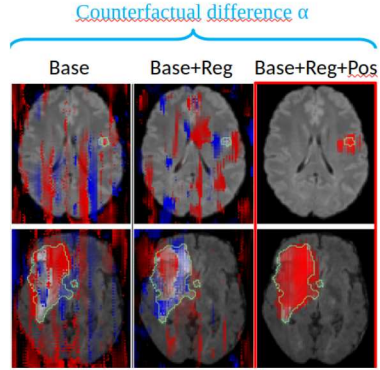
"gibbon"
99.3 % confidence

Goodfellow etal 2014

DLMI
Michaël Sdika
CREATIS - CNRS

Counterfactual explanation with a constrained architecture

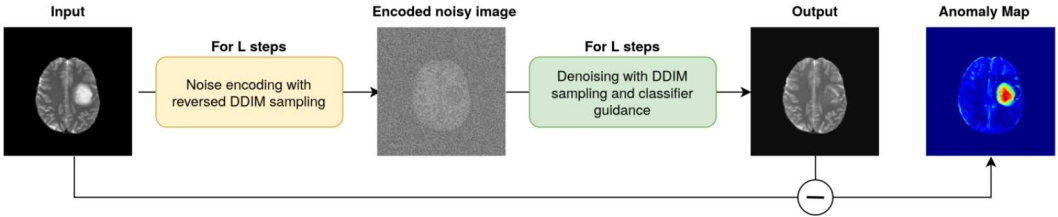




$$\min_{\alpha} M(f-\alpha) + \lambda \|\alpha\|_1$$

DLMI
Michaël Sdika
CREATIS - CNRS

Using Generative Models



counterfactual example s.t.:

- minimal change from input
- classif: opposite class
- output of a generative model

Wolleb etal MICCAI 2022, Diffusion Models for Medical Anomaly Detection

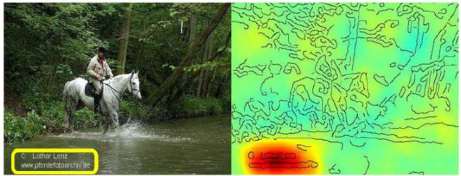
DLMI
Michaël Sdika
CREATIS - CNRS

Content

DLMI
Michaël Sdika
CREATIS - CNRS

Counfounder effect: Right for the wrong reason

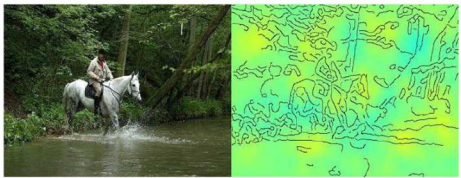
Horse-picture from Pascal VOC data set



Source tag present

↓

Classified as horse

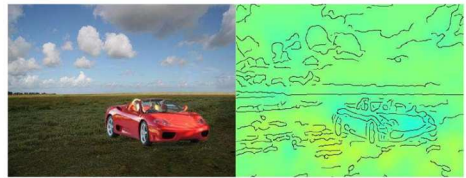


No source tag present

↓

Not classified as horse

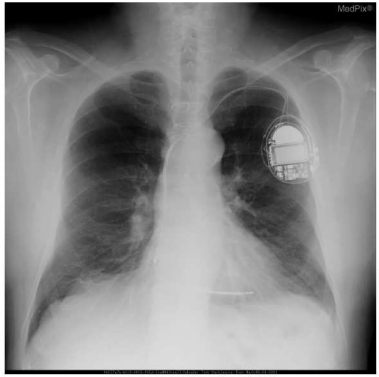
Artificial picture of a car



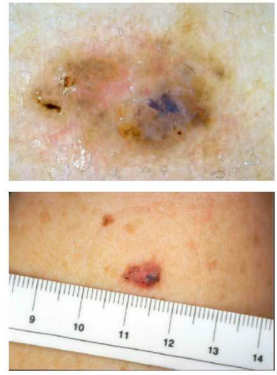
Lapuschkin et al. "Unmasking Clever Hans predictors and assessing what machines really learn." Nature communications 2019

DLMI
Michaël Sdika
CREATIS - CNRS

Obvious Counfounder in Medical Images



pacemaker on chest radiograph

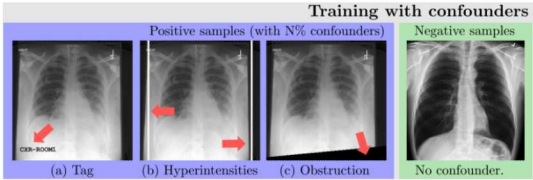


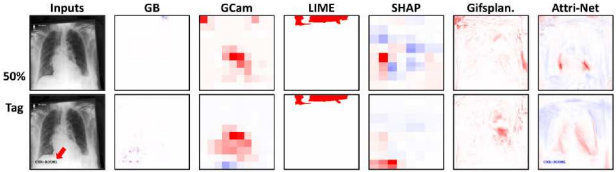
ruler on skin lesion images

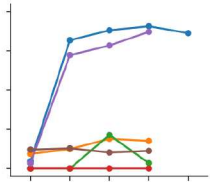
Kawahara et al. "Seven-point checklist and skin lesion classification using multitask multimodal neural nets." IEEE JBHI 22018.

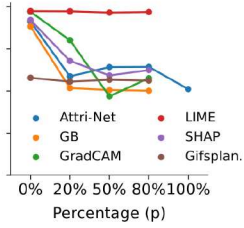
DLMI
Michaël Sdika
CREATIS - CNRS

Attribution maps comparison with artificial confounders









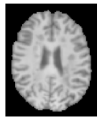
CS: % high att within confound area

NCC: pred. change $\Rightarrow$ att change

Sun, Koch, Baumgartner, MICCAI 2023, Right for the Wrong Reason: Can Interpretable ML Techniques Detect Spurious Correlations?

DLMI
Michaël Sdika
CREATIS - CNRS

The brain shape as a confounder



→ Classif MRI → P_{patho}

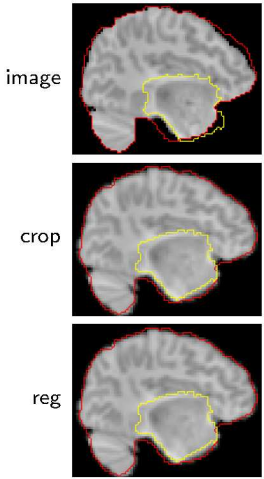
→ Classif Brain Mask → P_{patho}

	Classification task			TNR	TPR	BA	
Healthy vs Pathological	IXI	vs	BraTS	1.00	1.00	1.00 ✓	
	IXI	vs	OFSEP	0.54	0.70	0.62 ✓	
Healthy vs Healthy	IXI	vs	HCP	1.00	1.00	1.00 ✓	
	IXI	vs	IBC/kirby/MPI	1.00	0.86	0.93 ✓	
	HCP	vs	IBC/kirby/MPI	0.99	0.95	0.97 ✓	Age matching
Healthy vs Pathological	ADNI-CN	vs	ADNI-AD	0.35	0.74	0.55 ✗	Intra dataset

Wargnier-Dauchelle, Grenier, Sdika, MIDL 2024, An unexpected confounder: how brain shape can be used to classify MRI scans ?

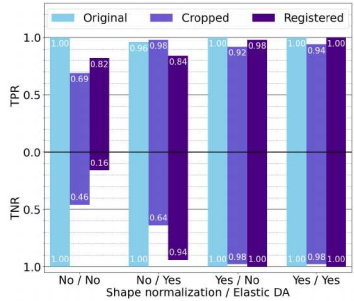
DLMI
Michaël Sdika
CREATIS - CNRS

The brain shape as a confounder: assessment and solution



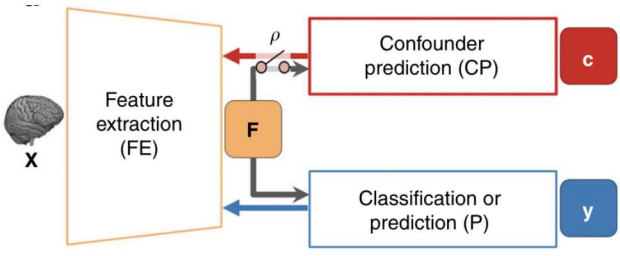
During test:
- Change an image from class A
  s.t. its brain shape is in class B
- Check if the performance decrease

Solution:
brain shape
normalization



DLMI
Michaël Sdika
CREATIS - CNRS

Confounder Free Training



CF-net architecture

- update FE+P
  estimate y

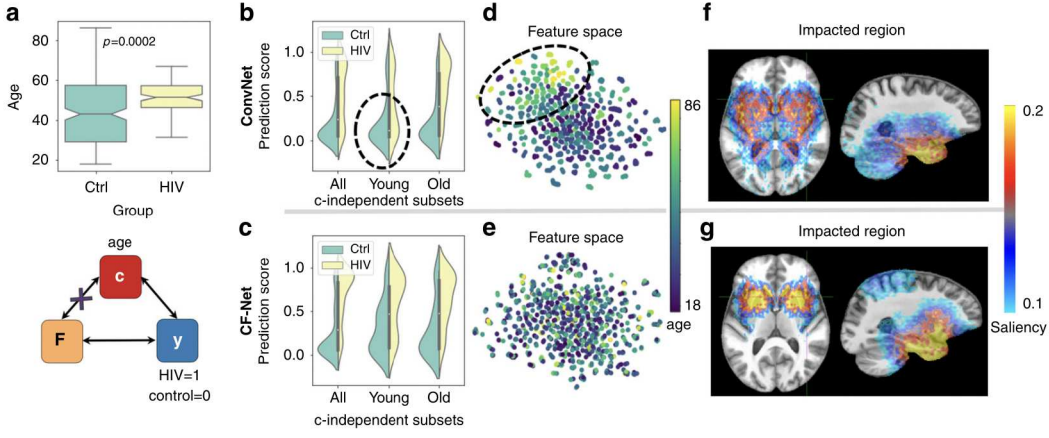
- update CP **on control only**
  estimate c
  avoid confounding of y on CP

- adversarial update FE:
  to fool CP
  $\Rightarrow$ F cannot predict c

Zhao et al, Nature Communications 2020, Training confounder-free deep learning models for medical applications

DLMI
Michaël Sdika
CREATIS - CNRS

Confounder Free Training: results



Zhao et al, Nature Communications 2020, Training confounder-free deep learning models for medical applications

Conclusion

DLMI
Michaël Sdika
CREATIS - CNRS

Thanks for your attention !!

DLMI
Michaël Sdika
CREATIS - CNRS

Path Methods: Integrated Gradient

$$A(I,B) = (I-B)\int_0^1 \frac{\partial F}{\partial I}\left(B+\alpha(I-B)\right)d\alpha$$

Completeness	$\sum_x A(I)_x = F(I) - F(B)$
Sensitivity	F do not depend on pixel $x \implies A(I)(x) = 0$
Linearity	$A(aF_1 + bF_2) = aA(F_1) + bA(F_2)$
Implementation	$\forall I,\ F_1(I) = F_2(I) \rightarrow A_1 = A_2$
Invariance	

Th. Only path methods always satisfy these 4 properties.

M Sundararajan, et al ICML 2017, Axiomatic attribution for deep networks

DLMI
Michaël Sdika
CREATIS - CNRS

Global vs Local Attributions

$$C = 1.05x_1 + 10x_2$$

$$x_1 = 1.000.000 \text{ and } x_2 = 10.000$$

Local:

$$R1 = \frac{\partial C}{\partial x_1} = 1.05$$

$$R2 = \frac{\partial C}{\partial x_2} = 10$$

How do you behave locally ?

Global:

$$R1 = x_1 \frac{\partial C}{\partial x_1} = 1.050.000$$

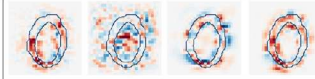
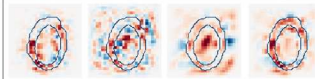
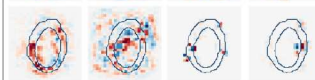
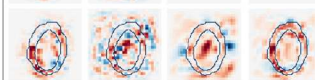
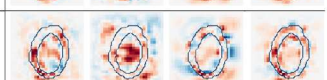
$$R2 = x_2 \frac{\partial C}{\partial x_2} = 10.000$$

How initial investments contributed ?

Ancona et al ICLR 2018, Towards better understanding of gradient-based attribution methods for deep neural networks

DLMI
Michaël Sdika
CREATIS - CNRS

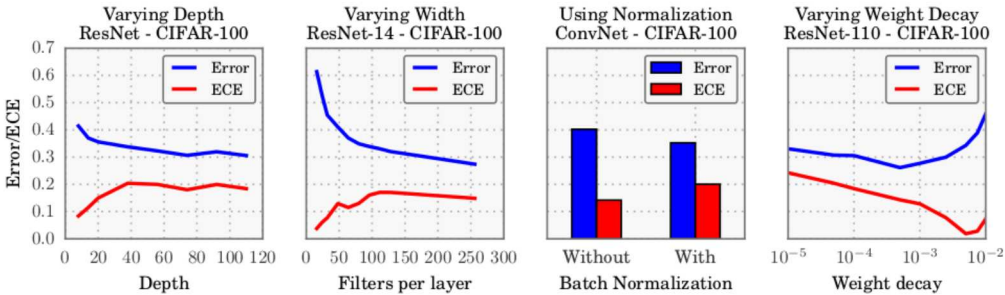
Retropagation rules: LRP, DeepLIFT, ...

Method	Attribution $R_i^c(x)$	Example of attributions on MNIST
		ReLU Tanh Sigmoid Softplus
Gradient * Input	$x_i \cdot \frac{\partial S_c(x)}{\partial x_i}$	
Integrated Gradient	$(x_i - \bar{x}_i) \cdot \int_{\alpha=0}^{1} \frac{\partial S_c(\tilde{x})}{\partial(\tilde{x}_i)}\Big\|_{\tilde{x}=\bar{x}+\alpha(x-\bar{x})} d\alpha$	
ϵ-LRP	$x_i \cdot \frac{\partial^g S_c(x)}{\partial x_i}, \quad g = \frac{f(z)}{z}$	
DeepLIFT	$(x_i - \bar{x}_i) \cdot \frac{\partial^g S_c(x)}{\partial x_i}, \quad g = \frac{f(z) - f(\bar{z})}{z - \bar{z}}$	
Occlusion-1	$S_c(x) - S_c(x_{[x_i=0]})$	

Ancona et al ICLR 2018, Towards better understanding of gradient-based attribution methods for deep neural networks

DLMI
Michaël Sdika
CREATIS - CNRS

Network probability output

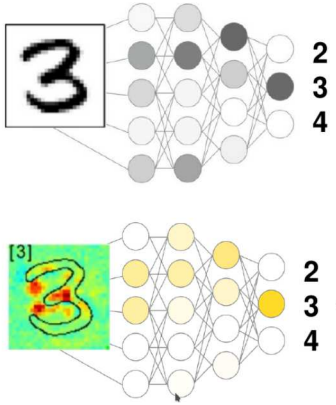


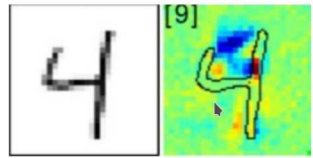
Guo *et. al.* 2017, On Calibration of Modern Neural Networks

About Interpreting Deep Learning...

DLMI
Michaël Sdika
CREATIS - CNRS

Layer-wise Relevance Propagation (LRP)

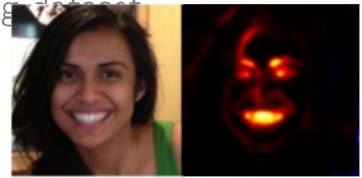




DLMI
Michaël Sdika
CREATIS - CNRS

Investigate bias on training dataset

- Age Predictions from face images
- with pretraining on ImageNet, laughing speaks against 60+
- model learned that old people do not laugh



- 25-32 years old relevance image

- 60+ years old relevance image (pre training on image net)

- 60+ years old relevance image (pre training on IMDB-WIKI)

MICCAI 2018 Tutorial, Explainable ML, Medical Applications, www.heatmapping.org
Lapuschkin *et. al.* Journal of Machine Learning Research 2016, The Layer-wise Relevance Propagation Toolbox for Artificial Neural Networks

DLMI
Michaël Sdika
CREATIS - CNRS

Investigate bias on training dataset

horse images in PASCAL VOC 2007

Image Fisher Vector CNN

Accuracy:	FV	80.45%
	CNN	81.60%

MICCAI 2018 Tutorial, Explainable ML, Medical Applications, www.heatmapping.org
Lapuschkin *et. al.* Journal of Machine Learning Research 2016, The Layer-wise Relevance Propagation Toolbox for Artificial Neural Networks

DLMI
Michaël Sdika
CREATIS - CNRS

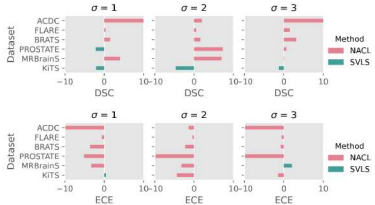
Spatially Varying Label Smoothing

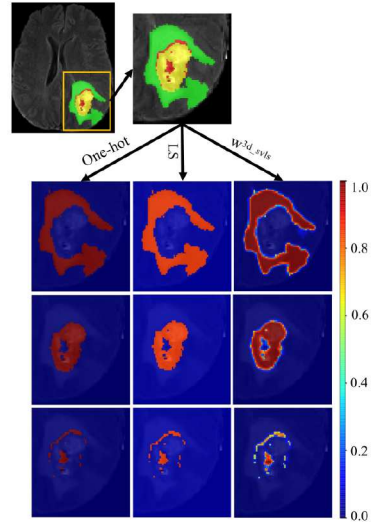
SVLS

- gaussian smooth GT label maps
- use it in CE

NACL

$$\min CE + P(\|l - \tau\|)$$





Islam & Glocker, IPML 2021. Spatially varying label smoothing: Capturing uncertainty from expert annotations
Murugesan ... Dolz , MEDIA 2024, Neighbor-Aware Calibration of Segmentation Networks with Penalty-Based Constraints

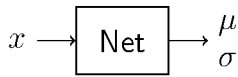
Regression

$$\text{MSE} = \|\hat{a} - a\|^2$$

Nix & Weigend, ICNN 1994. Estimating the mean and variance of the target probability distribution

DLMI
Michaël Sdika
CREATIS - CNRS

Regression



$\text{MSE} = \|\hat{a} - a\|^2$

$A(x) \sim \mathcal{N}(\mu(x), \sigma(x))$

likelyhood: $\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-\|\mu - a\|^2}{2\sigma^2}}$

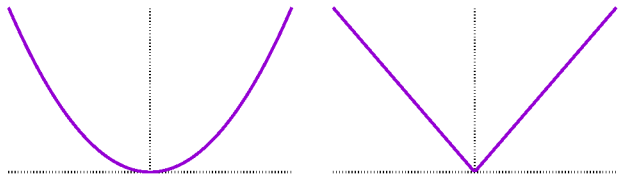
homoscedastic ($\sigma = \text{cst}$):

$\Rightarrow \text{MSE} = \text{NLL}$ $\longrightarrow$ NLL $: \log(\sigma) + \frac{\|\mu - a\|^2}{2\sigma^2}$

Nix & Weigend, ICNN 1994. Estimating the mean and variance of the target probability distribution

Quantile Regression

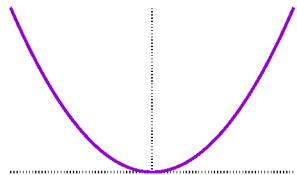
$$\min_x \sum_i r(x - x_i)$$

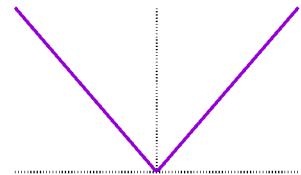


DLMI
Michaël Sdika
CREATIS - CNRS

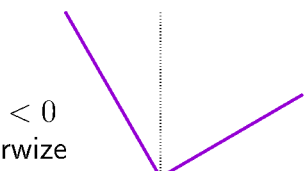
Quantile Regression

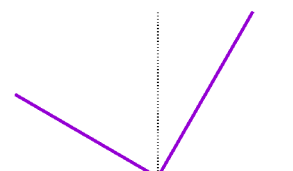
$$\min_x \sum_i r(x - x_i)$$





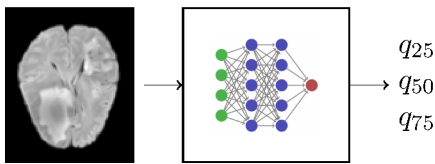
$$r_\tau(x) = \begin{cases} -(1-\tau)x & \text{if } x < 0 \\ x\tau & \text{otherwize} \end{cases}$$





DLMI
Michaël Sdika
CREATIS - CNRS

Quantile Regression



Loss: $r_{25}(q_{25} - y) + r_{50}(q_{50} - y) + r_{75}(q_{75} - y)$

DLMI
Michaël Sdika
CREATIS - CNRS

More on conformal prediction

Adaptive prediction set:

$$s(x,y) = 1 - \sum_{f_z(x) \geq f_y(x)} f_z(x)$$

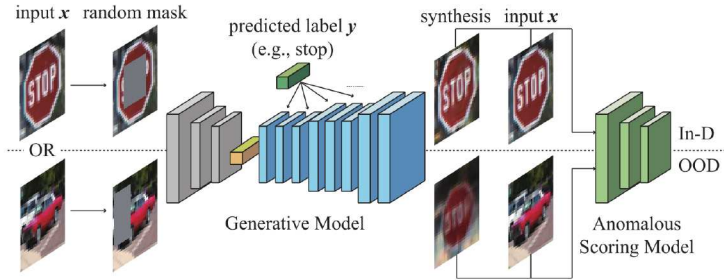
Regression:

- quantile regression: $\rightarrow f_q(x),\ f_{1-q}(x)$ (with $q = \alpha/2$)
- score: $s(x,y) = max\{f_q(x) - y, y - f_{1-q}(x)\}$
- ...

Angelopoulos & Bates, A gentle introduction to conformal prediction and distribution-free uncertainty quantification." arXiv 2021

DLMI
Michaël Sdika
CREATIS - CNRS

Reconstruction-based OOD



Denouden etal, arxiv 2018, Improving Reconstruction Autoencoder Out-of-distribution Detection with Mahalanobis Distance

DLMI
Michaël Sdika
CREATIS - CNRS

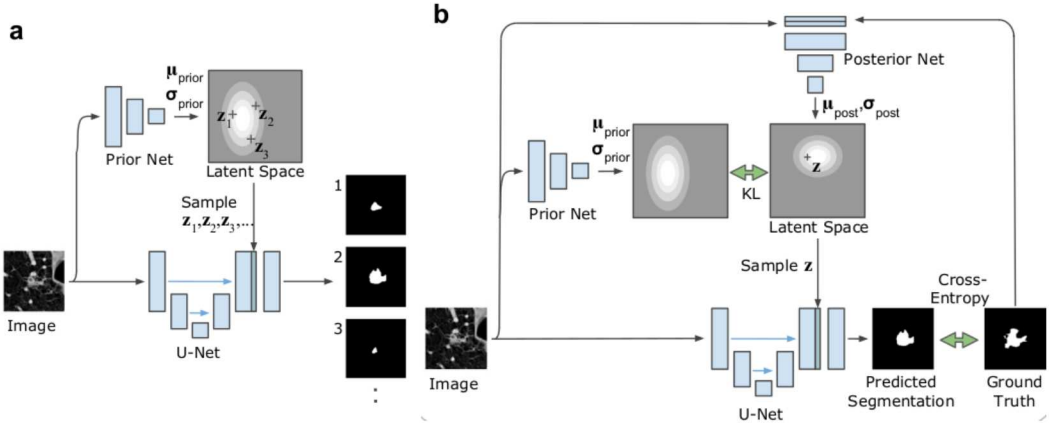
Uncertainty with respect to expert annotations

Lung CT images with segmentation made by 4 operators

Kohl *et. al.* NIPS 2018, A probabilistic u-net for segmentation of ambiguous images

DLMI
Michaël Sdika
CREATIS - CNRS

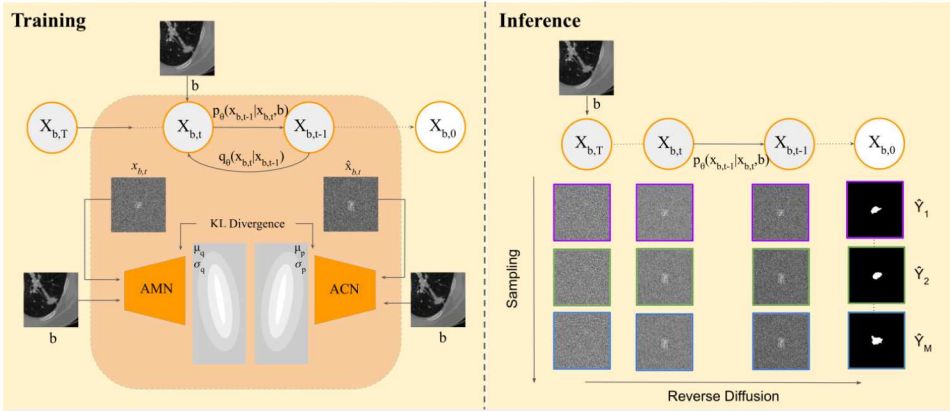
Uncertainty with respect to expert annotations



Kohl *et. al.* NIPS 2018, A probabilistic u-net for segmentation of ambiguous images

DLMI
Michaël Sdika
CREATIS - CNRS

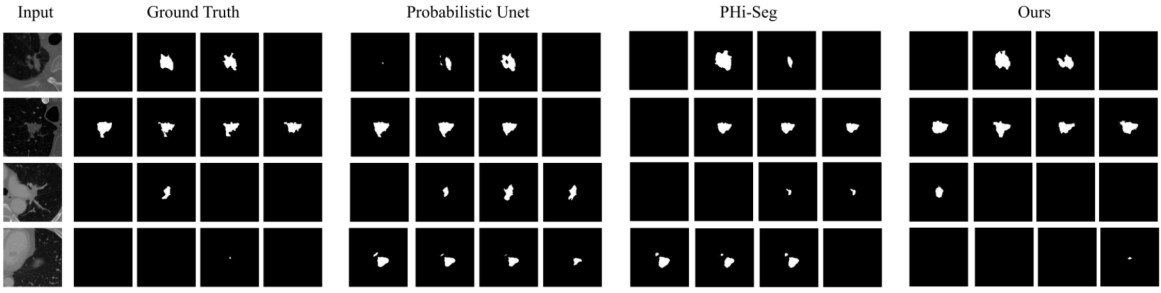
Generating segmentation distribution with diffusion models



Rahman etal, CVPR 2023, Ambiguous Medical Image Segmentation using Diffusion Models

DLMI
Michaël Sdika
CREATIS - CNRS

Generating segmentation distrbution with diffusion models



Method	GED (↓)	CI (↑)	D_{max} (↑)
DDPM-det-Seg [59]	1.081	0.616	0.548
DDPM-Prob-Seg	0.417	0.683	0.689
CIMD (Ours)	**0.321**	**0.759**	**0.915**

Rahman etal, CVPR 2023, Ambiguous Medical Image Segmentation using Diffusion Models