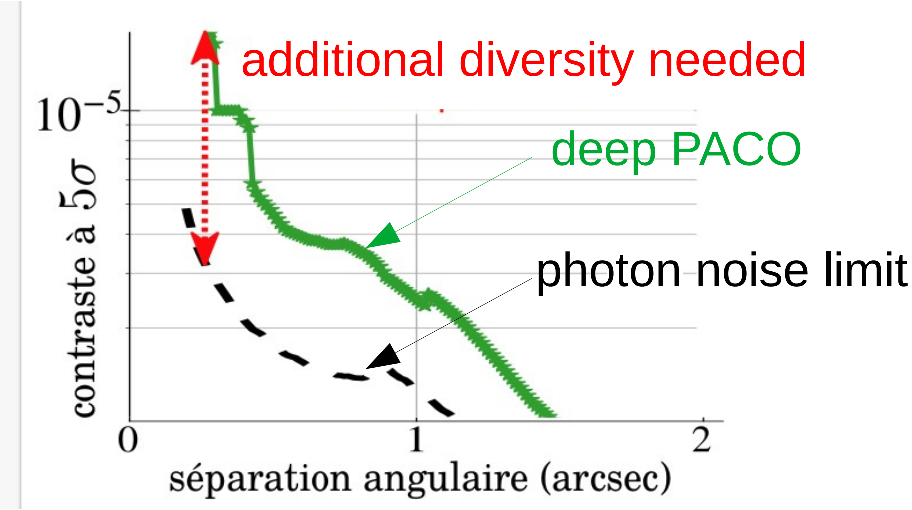
Brainstorming data science & AI

COBREX week - 04/10/2022

Some specifications for future detection / characterization algorithms

- building a model from several datasets
  → *see yesterday's discussions about RDI...*
  ...and today's discussions

(not only searching for image similarities!)



- deriving statistically grounded detection score & associated uncertainties

- including our knowledge about the problem (model-based approaches)

- including a form of joint detection/estimation
  → *see today's discussions*

- exploiting the metadata
  → *see today's discussions*

Agenda

- Beyond black-box approaches

  - control of the uncertainties (short review of review papers, Olivier, Théo = 10 min)

  - model-based learning (review of review papers, Théo, Olivier = 30+ min) → a focus on algorithm unrolling

  + learning a prior (plug and play approaches) on the nuisance component and/or on the objects of interest. T

  → The example of conventional imaging.

  - how to exploit these methods for your problems?: discussions & ideas (!ALL! = 20-30min)

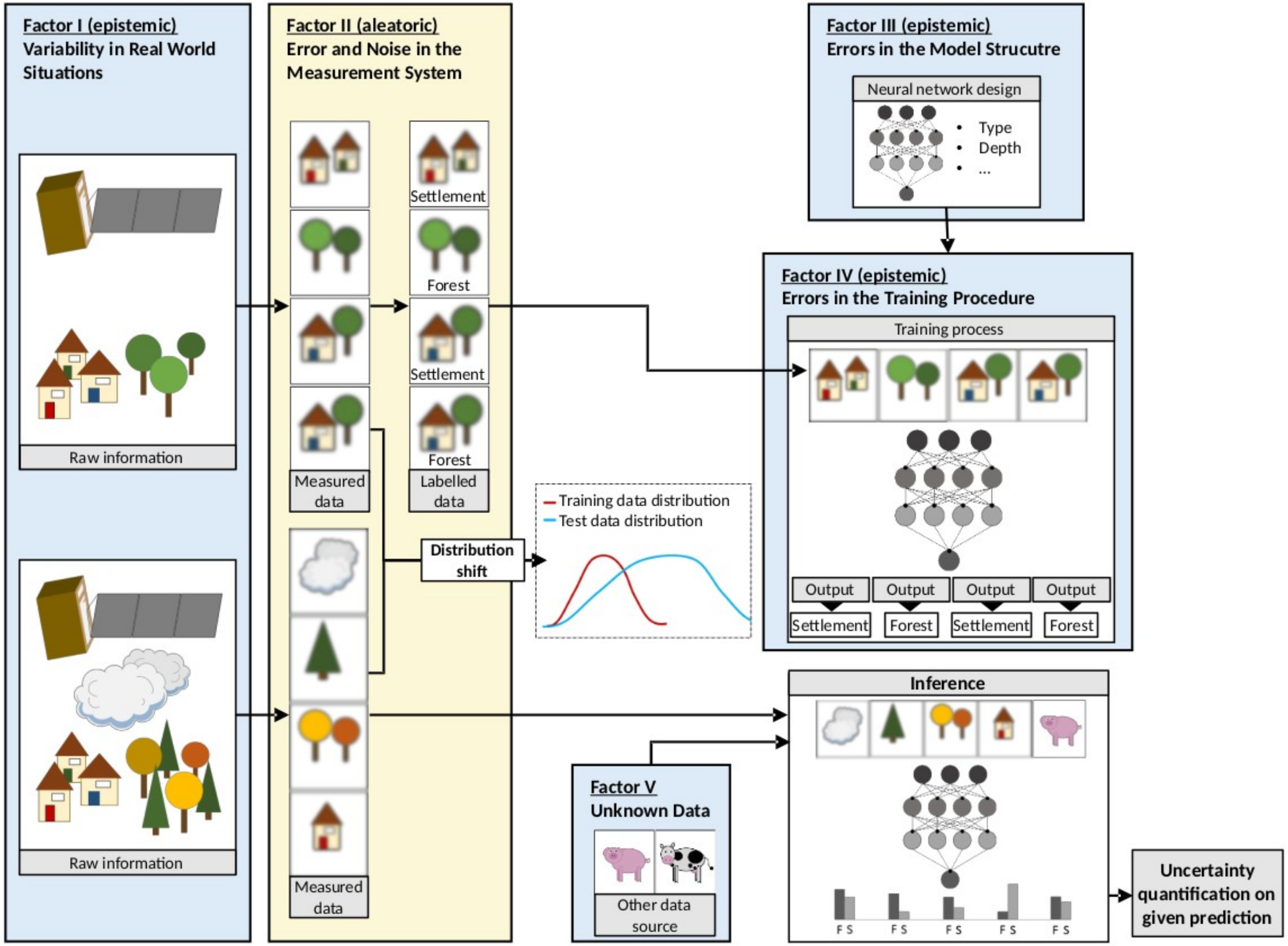
- Exploitation of the metadata

  - presentation of the available metadata (Julien = 15 min)

  - review of a post-processing algorithm for exoplanet detection & characterization exploiting metadata (Olivier = 15 min)

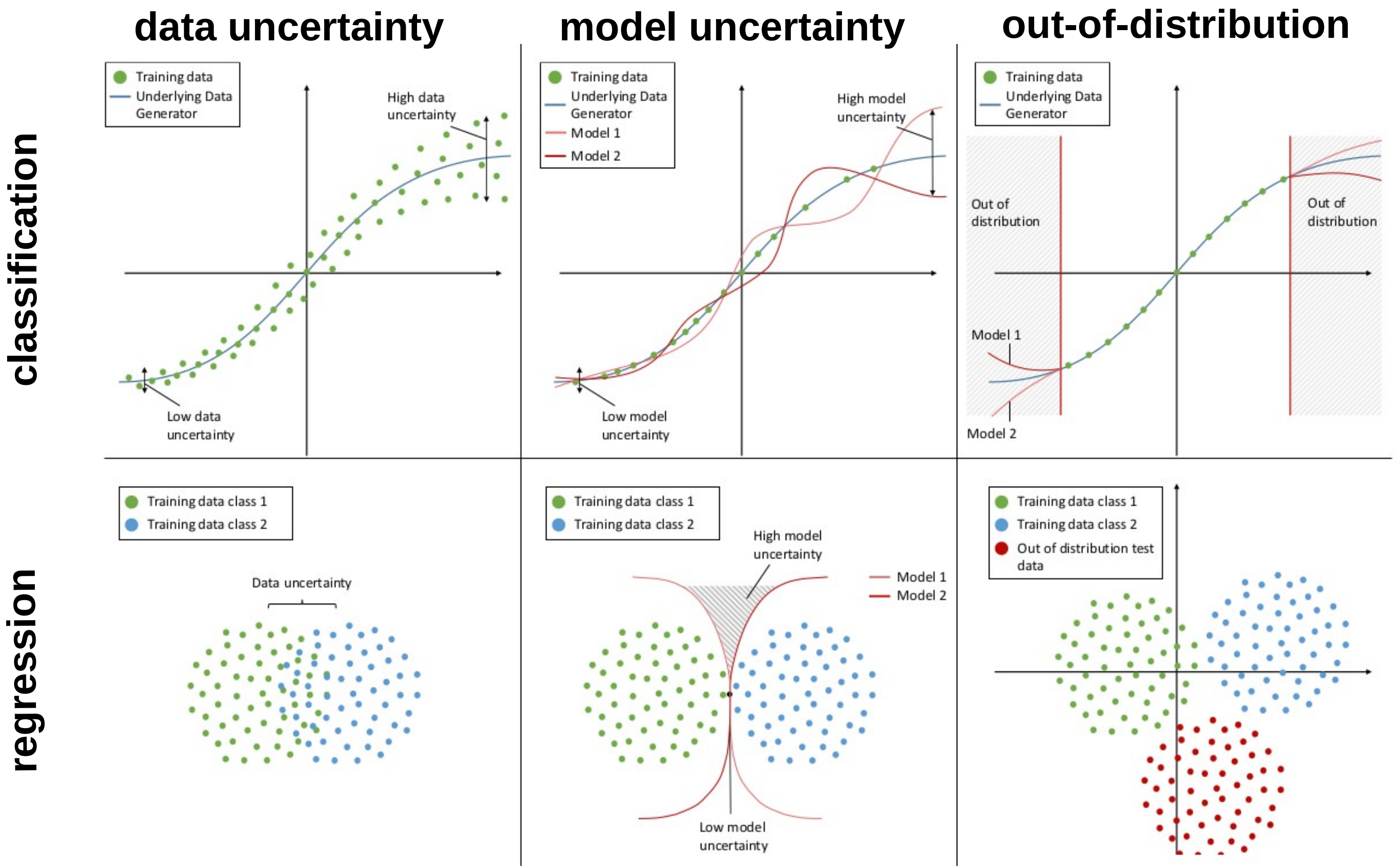
  - discussions & ideas (!ALL! = 30 min)

Different sources of uncertainties



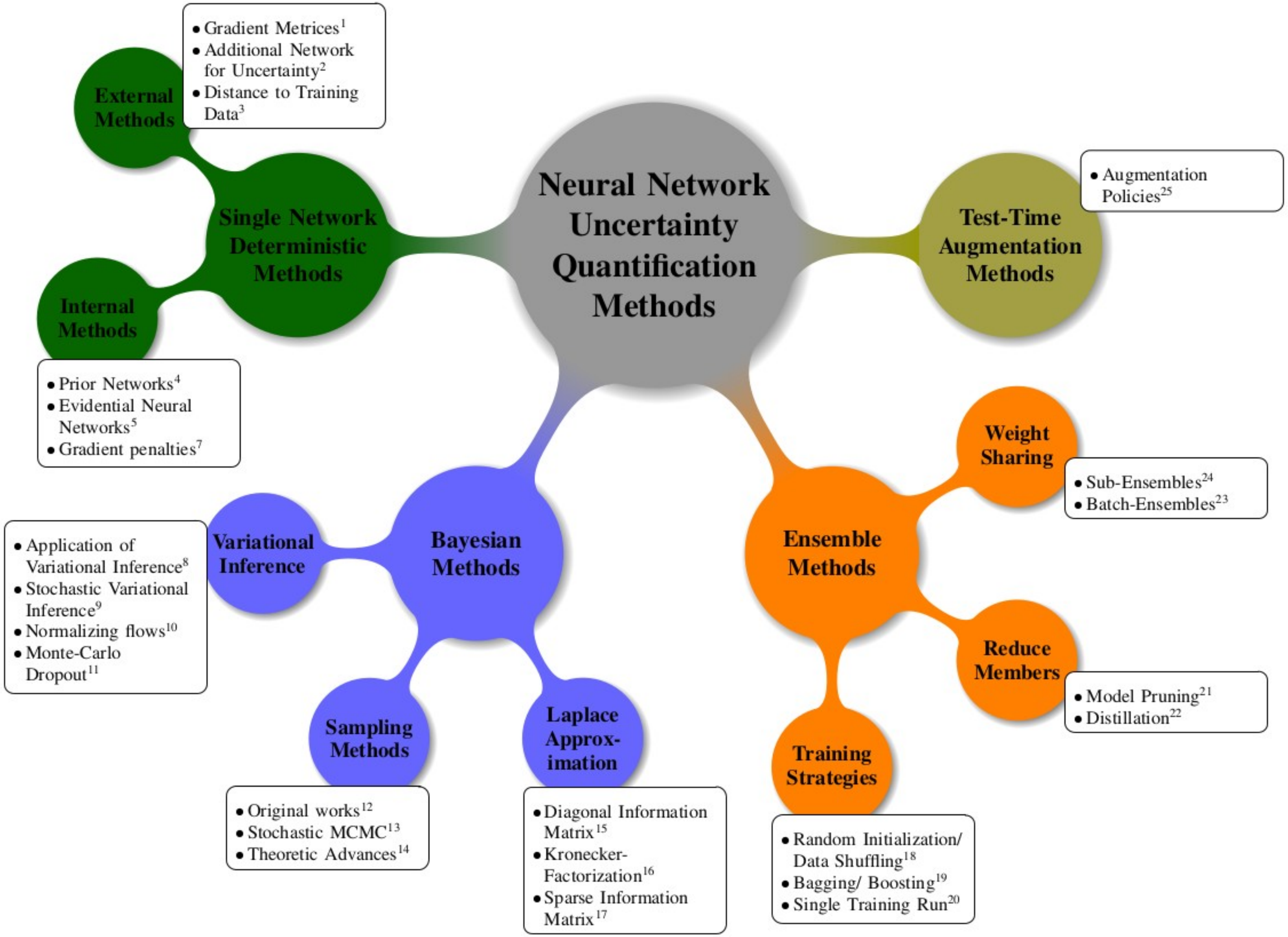
Gawlikowski+, « A Survey of Uncertainty in Deep Neural Networks », ACM Computing Surveys, 2022

Different sources of uncertainties



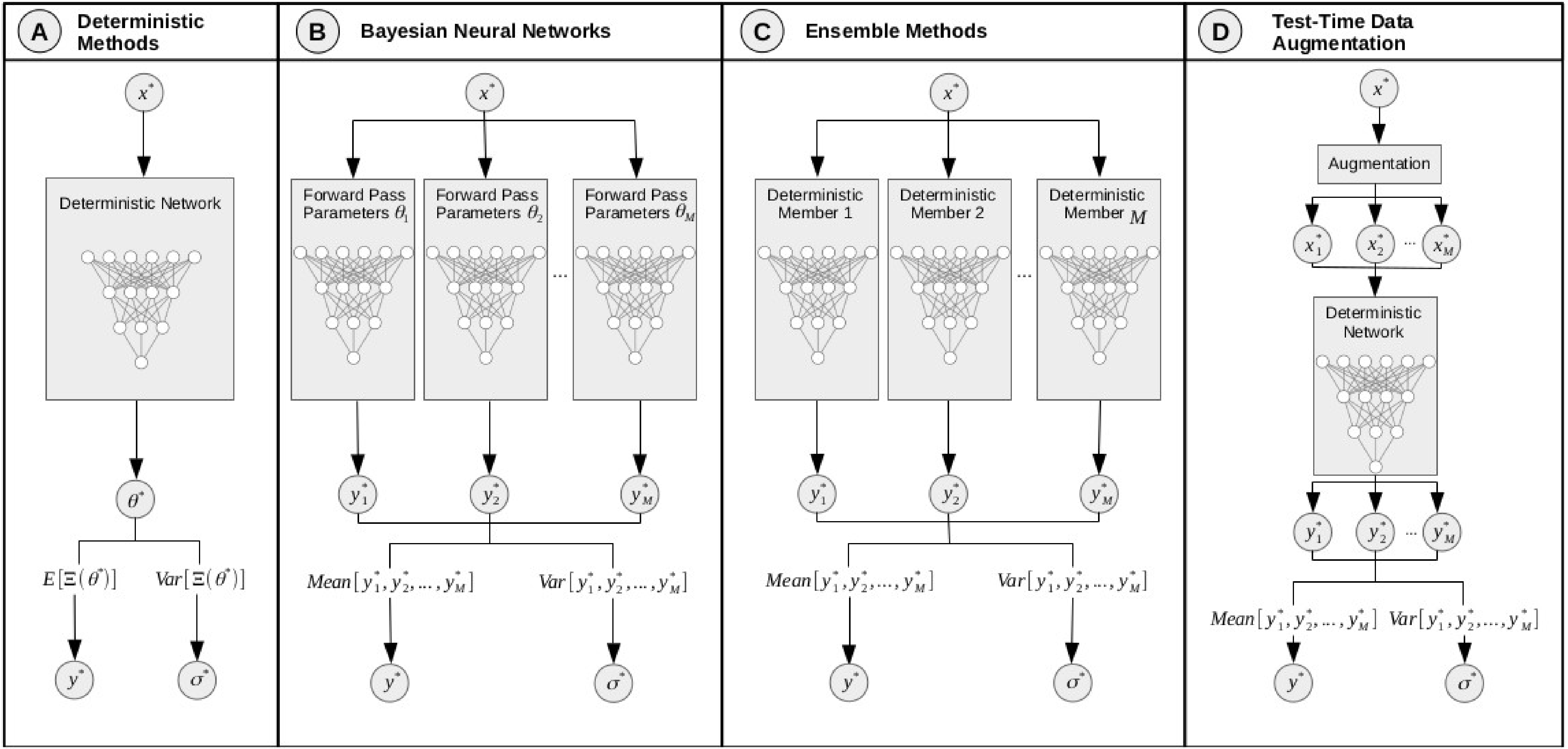
Gawlikowski+, « A Survey of Uncertainty in Deep Neural Networks », ACM Computing Surveys, 2022

Different types of methods to estimate the uncertainties



Gawlikowski+, « A Survey of Uncertainty in Deep Neural Networks », ACM Computing Surveys, 2022

Different types of methods to estimate the uncertainties



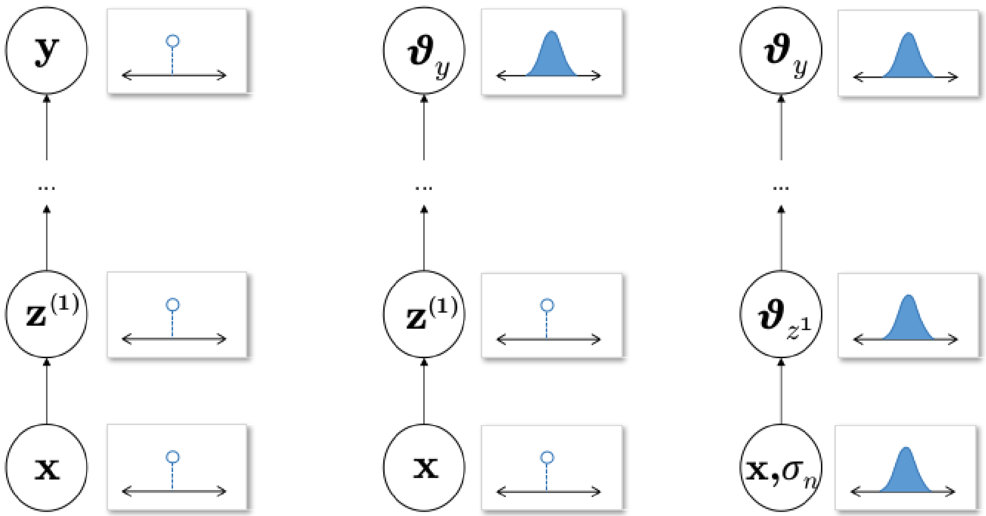
Gawlikowski+, « A Survey of Uncertainty in Deep Neural Networks », ACM Computing Surveys, 2022

Different types of methods to estimate the uncertainties

	Single Deterministic Networks	Bayesian Methods	Ensemble Methods	Test-Time Data Augmentation
Description	Approaches that receive an uncertainty quantification on a prediction of a deterministic neural network.	Model parameters are explicitly modeled as random variables. For a single forward pass the parameters are sampled from this distribution. Therefore, the prediction is stochastic and each prediction is based on different model weights.	The predictions of several models are combined into one prediction. A variety among the single models is crucial.	The prediction and uncertainty quantification at inference is based on several predictions resulting from different augmentations of the original input sample.
Description of Model Uncertainties	No	Yes	No	No
Need changes on existing networks	Depends on method	Yes	Yes (retrain several times)	No
Sensitivity to initialization and parameters of training process	High (in general)	Low (Usage of uninformative priors possible)	Low	Low
Number of networks trained	1	1	Several	1
Computational effort during training	Low	High	High	Low
Memory consumption during training	Low	Low	High	Low
Number of inputs per prediction	1	1	1	Several
Forward passes per prediction	1	Several	Several	Several
Evaluated modes	Single	Single	Multiple	Single
Computational effort during inference	Low (One forward pass, possibly some minor additional effort for uncertainty quantification)	High (sampling is either needed for explicit approach or for the approximation of intractable formulas)	High (Several models need to be evaluated)	High (Several augmentations and forward passes are performed)
Memory Consumption - Inference	Low	Low	High	Low

Gawlikowski+, « A Survey of Uncertainty in Deep Neural Networks », ACM Computing Surveys, 2022

Approach followed



(a) Determ. (b) ProbOut (c) ADF

Figure 1. **Uncertainties in CNNs**: (a) Traditional deep networks represent both activations and outputs as deterministic point estimates. (b) In this work, we explore the replacement of outputs by probabilistic output layers. (c) To go one step further, we also consider replacing all intermediate activations by distributions.

Approach followed

deep PACO: a variant for joint detection and estimation

Regression task: predicting images of flux $\widehat{\alpha}$ + confidence $\widehat{\boldsymbol{\sigma}}$

Sup. training: M triplets $\{\text{data}_m \text{ ; inj. locations } \boldsymbol{y}_m^{\text{GT}}, \text{ inj. flux } \boldsymbol{\alpha}_m^{\text{GT}}\}$

Link between detection and estimation: $\widehat{\text{SNR}}_m = \widehat{\boldsymbol{\alpha}}_m / \widehat{\boldsymbol{\sigma}}_m$ (N-pixels each)

$\Rightarrow$ *metric to control detection relevance (PFA, PD).*

Combined-loss for joint detection and estimation:

$$\ell = \sum_{m=1}^{M} \ell_{\text{det.}}(\widehat{\boldsymbol{y}}_m; \boldsymbol{y}_m^{\text{GT}}) + \lambda \ell_{\text{est.}}(\widehat{\boldsymbol{\alpha}}_m, \widehat{\boldsymbol{\sigma}}_m; \boldsymbol{\alpha}_m^{\text{GT}}) \quad \text{with:}$$

$$\ell_{\text{det.}} = \text{dice}(\widehat{\boldsymbol{y}}_m, \boldsymbol{y}_m^{\text{GT}}), \text{ st. } \widehat{\boldsymbol{y}}_m = \widehat{\text{SNR}}_m \geq \tau \text{ (thresholded SNR map)}$$

$$\boxed{\ell_{\text{est.}} = \sum_{n=1}^{N} \frac{(\widehat{\alpha}_{m,n} - \alpha_{m,n}^{\text{GT}})^2}{\widehat{\sigma}_{m,n}^2} + \log(\widehat{\sigma}_{m,n}) \text{ (max. lik. } \boldsymbol{\alpha} \sim \mathcal{N}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\sigma}}))}$$

Robustness wrt. τ: $\mathscr{L} = \sum_k w_k \ell_{\tau_k}$ s.t. $\sum_k w_k = 1$

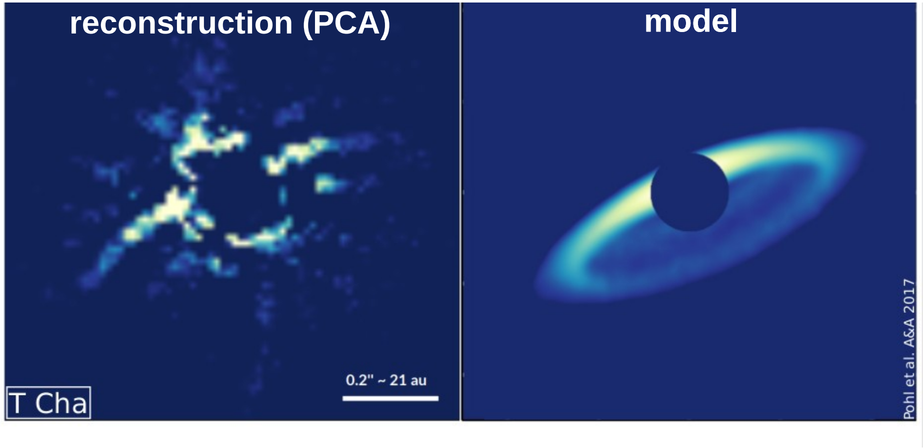
e.g. $\{(\tau_k, w_k)\} = \{(\tau_1 = 4, w_1 = 0.8), (\tau_2 = 2, w_2 = 0.1), (\tau_3 = 6, w_3 = 0.1)\}$.

Expected properties:

- joint detection/estimation,
  - *detection precision/recall rewarded by* $\ell_{\text{det.}}$
  - α *accuracy rewarded by* $\ell_{\text{est.}}$, σ *accuracy rewarded by* $\ell_{\text{dec.}}$,
  - σ *too conservative (large) penalized by* $\ell_{\text{dec.}}$ *and* $\ell_{\text{est.}}$.
- detection score statistically grounded,
- model can be built from several datasets.

Interaction massive data ⇔ physics-based simulations

exploit theoretical models to generate a priori information through Plug and Play methods



→ a way to improve the reconstructions by including some physics-based information at a macro level (smoothness, continuity etc.)

Open question: Inductive bias ?