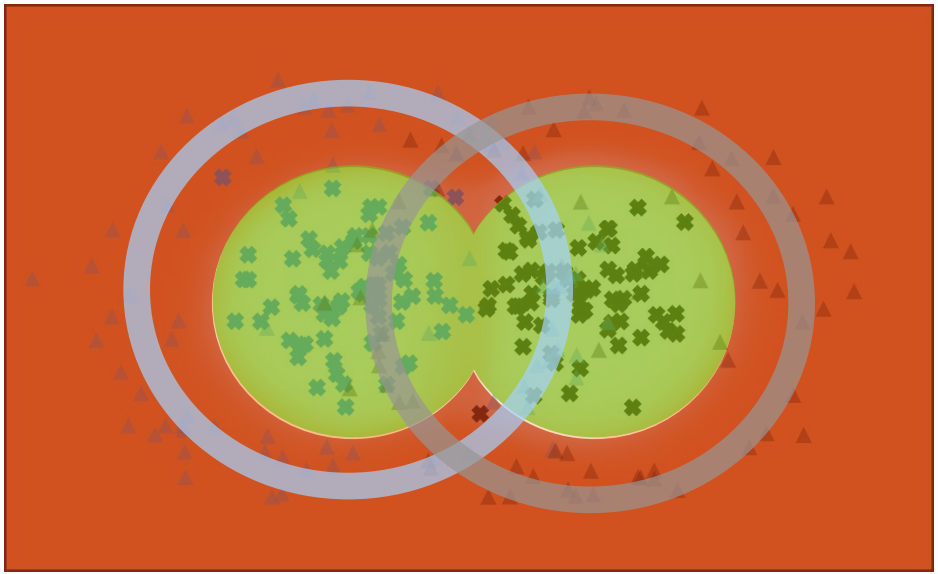
UQGAN: A Unified Model for Uncertainty Quantification of Deep Classifiers trained via Conditional GANs

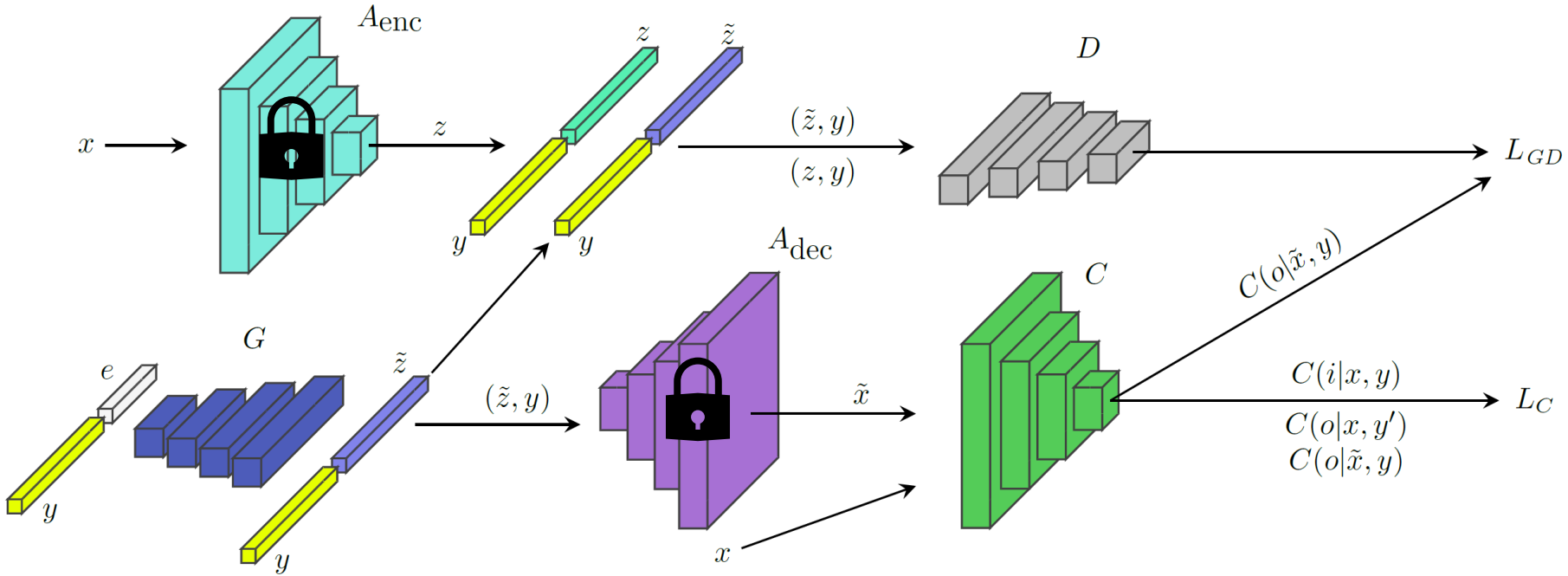
Philipp Oberdiek [1] Gernot Fink [1] Matthias Rottmann [2,3]

1: TU Dortmund University | **2:** University of Wuppertal | **3:** EPFL

Idea



Architecture



Results (Tiny ImageNet)

Method	In-Distribution			Out-of-Distribution			
	Accuracy ↑	AUROC S/F ↑	ECE ↓	AUROC ↑	AUPR-In ↑	AUPR-Out ↑	FPR@95% TPR ↓
Ours	34.28 (0.37)	71.90 (0.47)	48.94 (0.66)	79.25 (1.61)	26.35 (2.25)	97.66 (0.20)	47.22 (1.85)
Ours with MC-Dropout	**45.60 (0.43)**	79.18 (0.42)	5.92 (0.38)	**94.96 (0.13)**	**59.76 (0.64)**	**99.51 (0.01)**	**13.72 (0.30)**
One-vs-All Baseline	35.18 (0.26)	76.23 (0.43)	11.59 (4.41)	55.19 (2.29)	17.97 (2.41)	90.97 (0.69)	97.32 (0.62)
Max. Softmax [15]	36.06 (0.30)	78.56 (0.68)	26.01 (7.25)	61.53 (1.04)	21.07 (0.98)	93.29 (0.29)	92.51 (0.87)
Entropy	36.06 (0.30)	79.39 (0.74)	26.01 (7.25)	62.44 (1.31)	21.90 (0.86)	93.16 (0.46)	93.79 (1.28)
Bayes-by-Backprop [2]	32.31 (0.43)	78.44 (0.91)	19.39 (0.62)	68.05 (2.29)	21.24 (2.14)	95.23 (0.52)	81.70 (3.53)
MC-Dropout [7]	43.48 (0.53)	80.63 (0.30)	**2.79 (0.35)**	63.35 (4.23)	27.11 (2.29)	92.78 (1.06)	95.86 (1.28)
Deep-Ensembles [24]	42.48 (0.22)	**81.29 (0.38)**	16.50 (6.94)	67.76 (0.27)	30.85 (0.42)	93.93 (0.07)	93.79 (0.23)
Confident Classifier [27]	36.07 (0.53)	78.66 (0.67)	35.48 (2.13)	59.99 (1.84)	20.58 (1.61)	92.66 (0.39)	94.49 (0.51)
GEN [40]	30.54 (0.84)	73.40 (0.91)	28.79 (0.79)	84.65 (5.42)	36.86 (8.25)	98.28 (0.69)	39.67 (12.50)
Entropy Oracle	37.18 (0.50)	79.50 (0.63)	17.05 (2.82)	85.43 (1.90)	41.95 (2.31)	98.27 (0.29)	48.74 (6.63)
One-vs-All Oracle	34.92 (0.56)	75.10 (0.92)	20.61 (3.50)	95.75 (0.09)	59.69 (0.50)	99.59 (0.01)	10.30 (0.46)

Summary

- GAN-based model yielding a classifier with complete uncertainty quantification
- Model allows to distinguish between aleatoric and epistemic uncertainty
- Conditional GAN is trained with a Wasserstein-based loss function in the latent space of a conditional autoencoder
  ⟶ Reduces influence of adversarial directions
- Novel low-dimensional regularizer for improved class shielding
- Improve over the OoD detection and FP detection performance of SotA GAN-training based classifiers

Thank you for your attention

Code: https://github.com/RonMcKay/UQGAN

Contact: philipp.oberdiek@cs.tu-dortmund.de