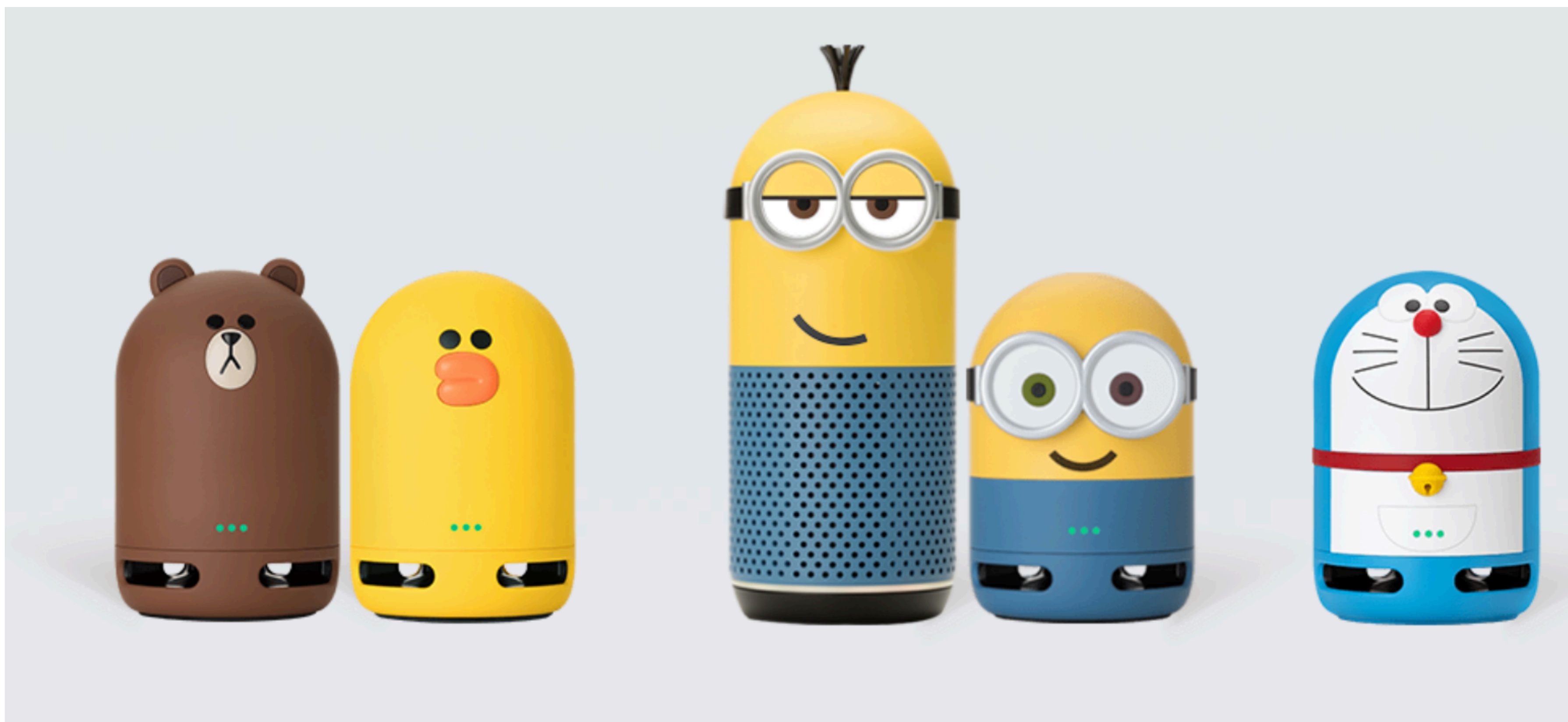
An Empirical Evaluation on the Generalization Ability of Regularization Methods

Sanghyuk Chun, CLOVA AI Research

NAVER | LINE Clova

Machine learning is finally working!



Clova

AI Speaker (Clova Friends)

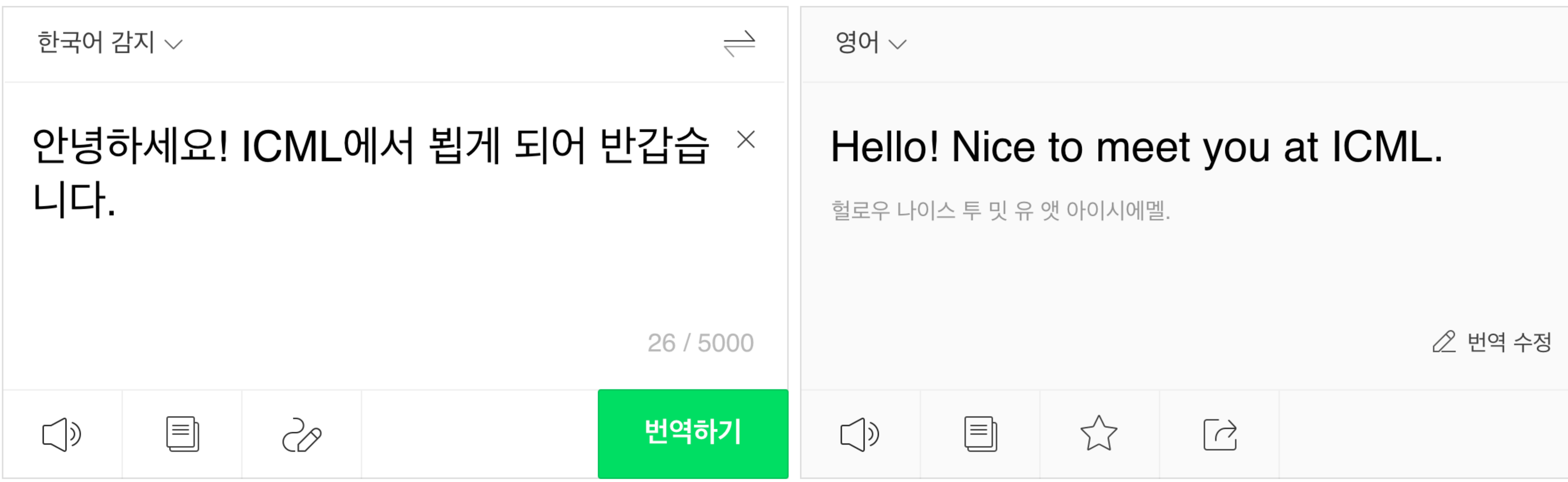
Speech Recognition

Speech Synthesis

Natural Language Understanding

Retrieval, Recommendation, ...

Machine learning is finally working!



papago

Papago
End-to-end Machine Translation

Machine learning is finally working!

築技
TSUKI
WAZA
築地の軒先で売られているのは、
ものだけじゃないんです。
開場から80年以上の歴史で育まれた
目利
きの知恵やノウバウを
お客さまに伝え
持ち帰っていただく。
だから通えば通うほど、
料理の腕が上がっていく。
築地はそんな市場です。
すべての店に必ずある築地の技「築技(つきわざ)」。
来る開場
100年目
さらに先の未来に向けて
磨き続けていきますので
どうぞこれからも築地にお越しください。

Clova OCR
Text Detection
Text Recognition
Document Parsing

Machine learning is finally working!

築技
TSUKI
WAZA
築地の軒先で売られているのは、
ものだけじゃないんです。
開場から80年以上の歴史で育まれた
目利
きの知恵やノウハウを
お客さまに伝え
持ち帰っていただく。
だから通えば通うほど、
料理の腕が上がっていく。
築地はそんな市場です。
すべての店に必ずある築地の技「築技(つきわざ)」。
来る開場
100年目
さらに先の未来に向けて
磨き続けていきますので
どうぞこれからも築地にお越しください。

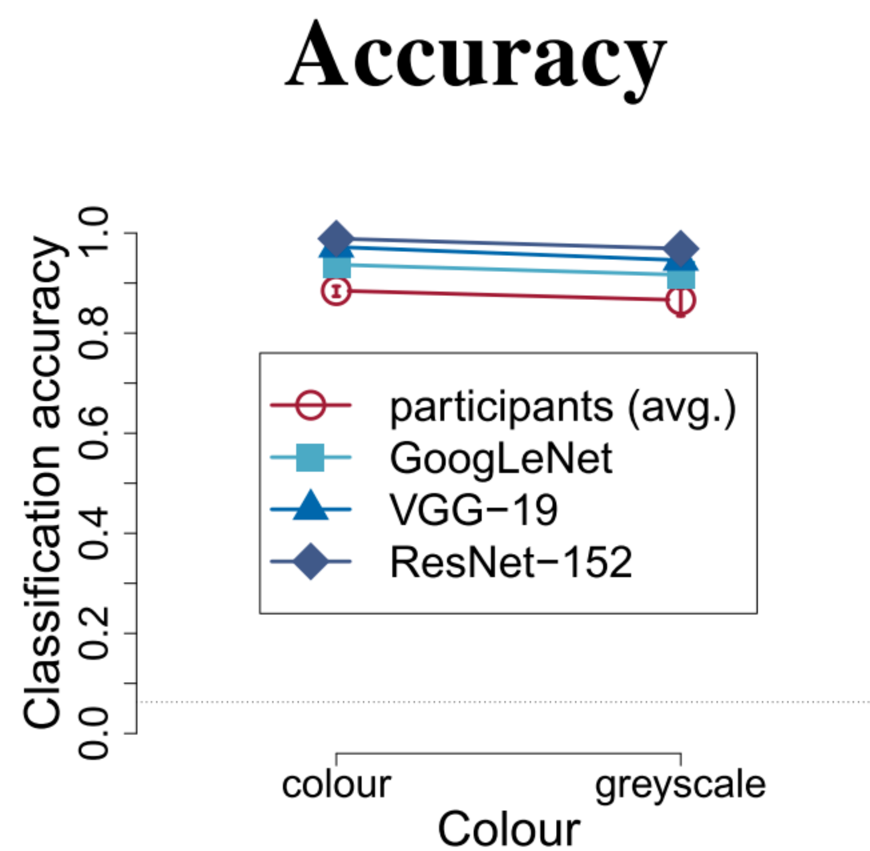
Clova OCR
Text Detection
Text Recognition
Document Parsing

Our new paper "Character Region Awareness for Text Detection" will be appeared in this CVPR!
(Thu, June 20, 2019 10 AM, #4706)

Human-level performance by ML models.

	Top-5 error
GoogLeNet	6.8%
ResNet	3.6%
Human (Andrej Karpathy)	5.1%

Deep models outperform humans in ImageNet validation top-5



Human vs. Deep models in selected ImageNet classes

Andrej Karpathy. http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/
Geirhos, Robert, et al. "Generalisation in humans and deep neural networks." Advances in Neural Information Processing Systems. 2018.

Human-level performance by ML models.

System	Dev EM	Dev F1	Test EM	Test F1
Top Leaderboard Systems (Dec 10th, 2018)				
Human	-	-	82.3	91.2
#1 Ensemble - nlnet	-	-	86.0	91.7
#2 Ensemble - QANet	-	-	84.5	90.5
Published				
BiDAF+ELMo (Single)	-	85.6	-	85.8
R.M. Reader (Ensemble)	81.2	87.9	82.3	88.5
Ours				
$\text{BERT}_{\text{BASE}}$ (Single)	80.8	88.5	-	-
$\text{BERT}_{\text{LARGE}}$ (Single)	84.1	90.9	-	-
$\text{BERT}_{\text{LARGE}}$ (Ensemble)	85.8	91.8	-	-
$\text{BERT}_{\text{LARGE}}$ (Sgl.+TriviaQA)	**84.2**	**91.1**	**85.1**	**91.8**
$\text{BERT}_{\text{LARGE}}$ (Ens.+TriviaQA)	**86.2**	**92.2**	**87.4**	**93.2**

Table 2: SQuAD 1.1 results. The BERT ensemble is 7x systems which use different pre-training checkpoints and fine-tuning seeds.

System	Dev EM	Dev F1	Test EM	Test F1
Top Leaderboard Systems (Dec 10th, 2018)				
Human	86.3	89.0	86.9	89.5
#1 Single - MIR-MRC (F-Net)	-	-	74.8	78.0
#2 Single - nlnet	-	-	74.2	77.1
Published				
unet (Ensemble)	-	-	71.4	74.9
SLQA+ (Single)	-		71.4	74.4
Ours				
$\text{BERT}_{\text{LARGE}}$ (Single)	78.7	81.9	80.0	83.1

Table 3: SQuAD 2.0 results. We exclude entries that use BERT as one of their components.

Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

Human-level performance by ML models.

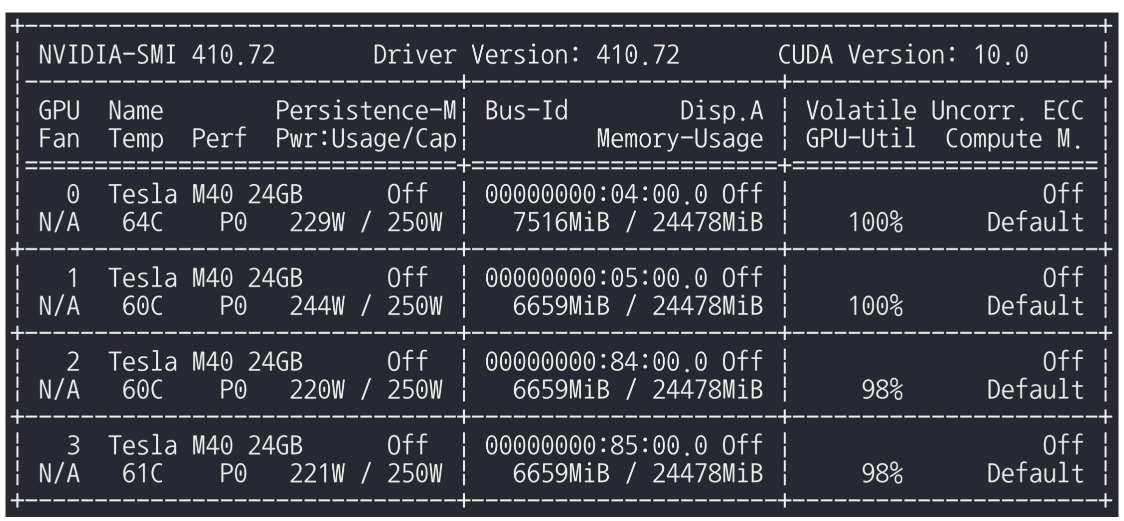
	Open-Ended				Multiple-Choice			
	All	Y/N	Num.	Other	All	Y/N	Num.	Other
DPPnet [19]	57.36	80.28	36.92	42.24	62.69	80.35	38.79	52.79
D-NMN [2]	58.00	-	-	-	-	-	-	-
Deep Q+I [15]	58.16	80.56	36.53	43.73	63.09	80.59	37.70	53.64
SAN [29]	58.90	-	-	-	-	-	-	-
ACK [27]	59.44	81.07	37.12	45.83	-	-	-	-
FDA [8]	59.54	81.34	35.67	46.10	64.18	81.25	38.30	55.20
DMN+ [28]	60.36	80.43	36.82	48.33	-	-	-	-
MRN	**61.84**	**82.39**	**38.23**	**49.41**	**66.33**	**82.41**	**39.57**	**58.40**
Human [1]	83.30	95.77	83.39	72.67	-	-	-	-

Kim, Jin-Hwa, et al. "Multimodal residual learning for visual qa." Advances in neural information processing systems. 2016.

Question:
ML models are perfect?

Question:
ML models are perfect?
... So can we just leave
the office?

ML models are not perfect



```
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 410.72       Driver Version: 410.72       CUDA Version: 10.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla M40 24GB      Off  | 00000000:04:00.0 Off |                  Off |
| N/A   64C    P0   229W / 250W |   7516MiB / 24478MiB |    100%      Default |
+-------------------------------+----------------------+----------------------+
|   1  Tesla M40 24GB      Off  | 00000000:05:00.0 Off |                  Off |
| N/A   60C    P0   244W / 250W |   6659MiB / 24478MiB |    100%      Default |
+-------------------------------+----------------------+----------------------+
|   2  Tesla M40 24GB      Off  | 00000000:84:00.0 Off |                  Off |
| N/A   60C    P0   220W / 250W |   6659MiB / 24478MiB |     98%      Default |
+-------------------------------+----------------------+----------------------+
|   3  Tesla M40 24GB      Off  | 00000000:85:00.0 Off |                  Off |
| N/A   61C    P0   221W / 250W |   6659MiB / 24478MiB |     98%      Default |
+-------------------------------+----------------------+----------------------+
```

Expensive

https://www.youtube.com/watch?v=MIbFvK2S9g8

Unreliable

Thys, Simen, et al. "Fooling automated surveillance cameras: adversarial patches to attack person detection." CVPR 2019

What are the issues?

Expensive

Supervision:
Human labeling

Heavy resources
for training

Heavy resources
for inference

What are the issues?

Expensive	Supervision: Human labeling	Heavy resources for training	Heavy resources for inference
	Weakly/semi-supervised ML	Smart res. alloc. sys. (NSML) / AutoML	Lightweight model for CPU inference

What are the issues?

Expensive	Supervision: Human labeling	Heavy resources for training	Heavy resources for inference
	Weakly/semi-supervised ML	Smart res. alloc. sys. (NSML) / AutoML	Lightweight model for CPU inference
Unreliable	Too confident	Not robust	

What are the issues?

Expensive

Supervision: Human labeling	Weakly/semi-supervised ML
Heavy resources for training	Smart res. alloc. sys. (NSML) / AutoML
Heavy resources for inference	Lightweight model for CPU inference

Unreliable

Too confident	Bayesian / probabilistic ML
Not robust	ML robustness

What are the issues?

Expensive

Supervision: Human labeling

Weakly/semi-supervised ML

Heavy resources

Not covered today (but we're interested in!)

(NSML) / AutoML

Heavy resources for inference

Lightweight model for CPU inference

Unreliable

Too confident

Bayesian / probabilistic ML

Not robust

ML robustness

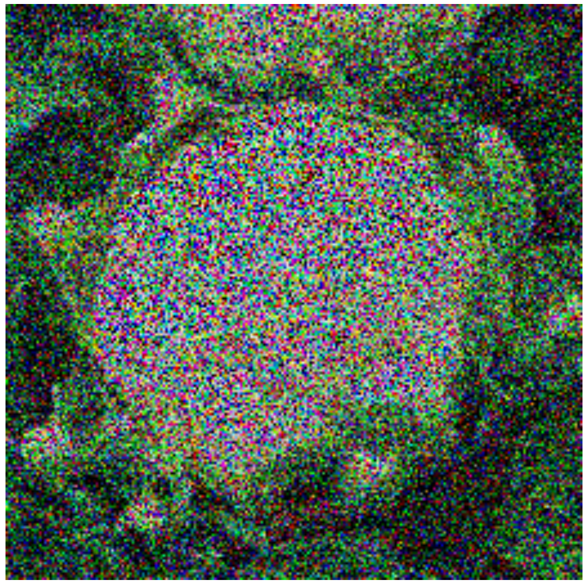
For the Remaining Talk,

- Introduction to ML robustness and uncertainty estimates
- Unexpected improvements of robustness & uncertainty by state-of-the-art regularization techniques
- Side topic: robustness in non-vision data (music)

DNNs behave fundamentally differently from humans.

Cauliflower (1.0)
(Clean Image)

brain coral (1.0)
(adversarially attacked)



bubble (0.5)
(+ Gaussian Noise)

Digital clock (0.2)
(Out-of-dist.)

DNNs are easily fooled.

Cauliflower (1.0)
(Clean Image)

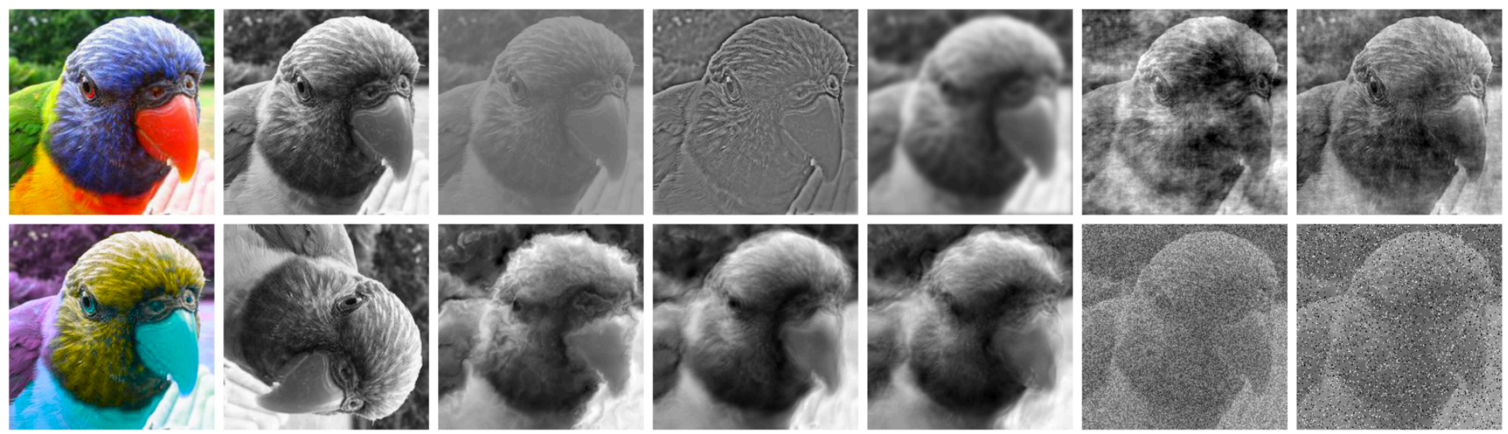
+

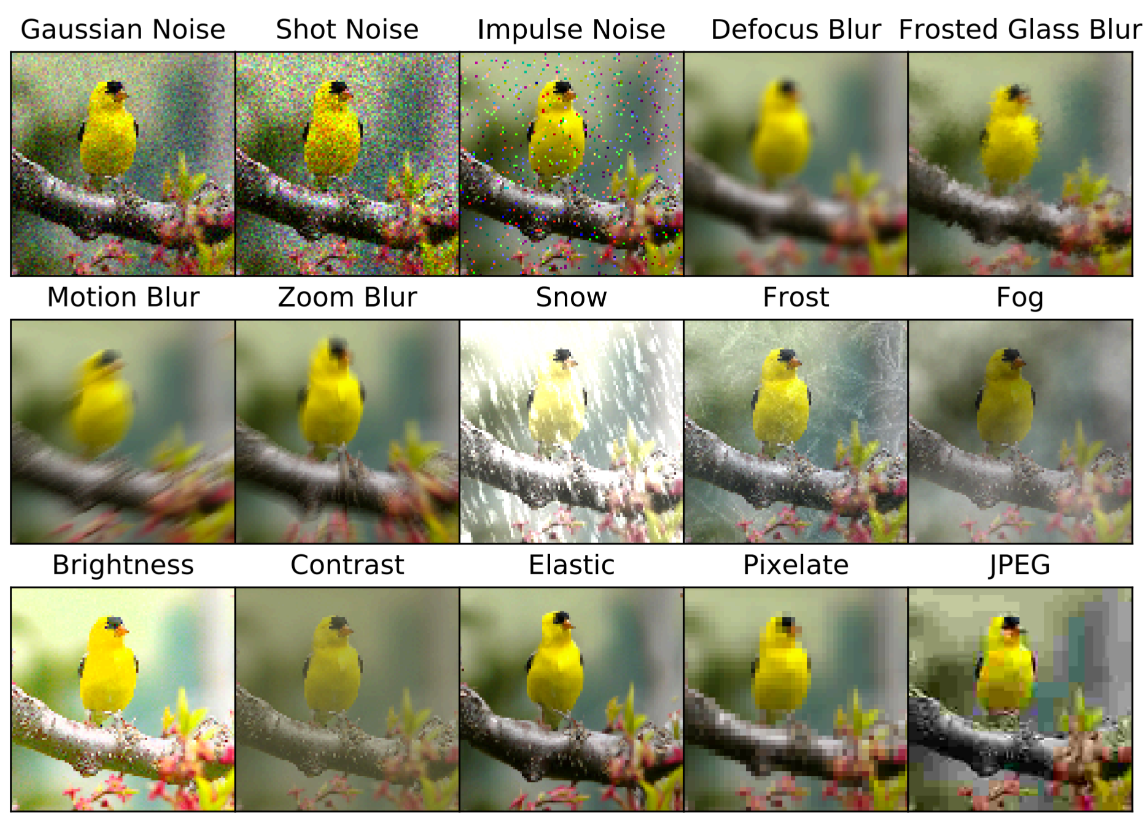
Human
imperceptible noise

=

brain coral (1.0)
(adversarially attacked)

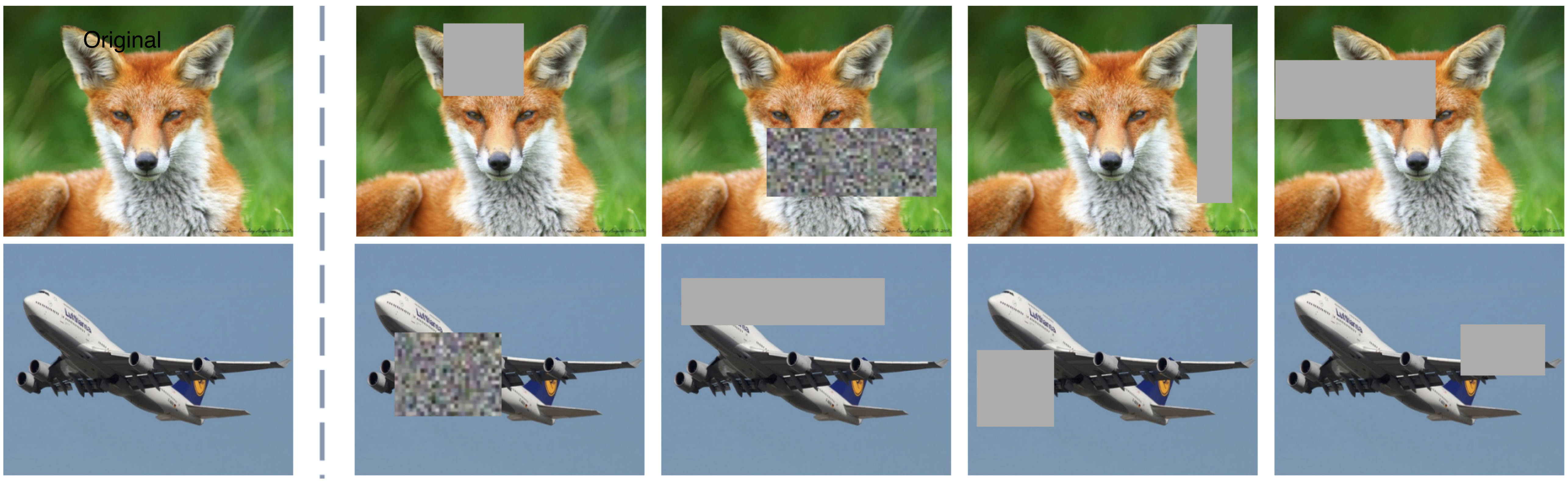
DNNs are unstable against natural corruptions.





Geirhos, Robert, et al. "Generalisation in humans and deep neural networks." Advances in Neural Information Processing Systems. 2018.
Hendrycks, Dan, and Thomas Dietterich. "Benchmarking neural network robustness to common corruptions and perturbations." ICLR 2019

Random erasing to improve occlusion stability.



Zhong, Zhun, et al. "Random erasing data augmentation." arXiv preprint arXiv:1708.04896 (2017).

CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features.

Sangdoo Yun
Clova AI Research, Naver Corp.

Dongyoon Han
Clova AI Research, Naver Corp.

Seong Joon Oh
Clova AI Research, LINE Plus Corp.

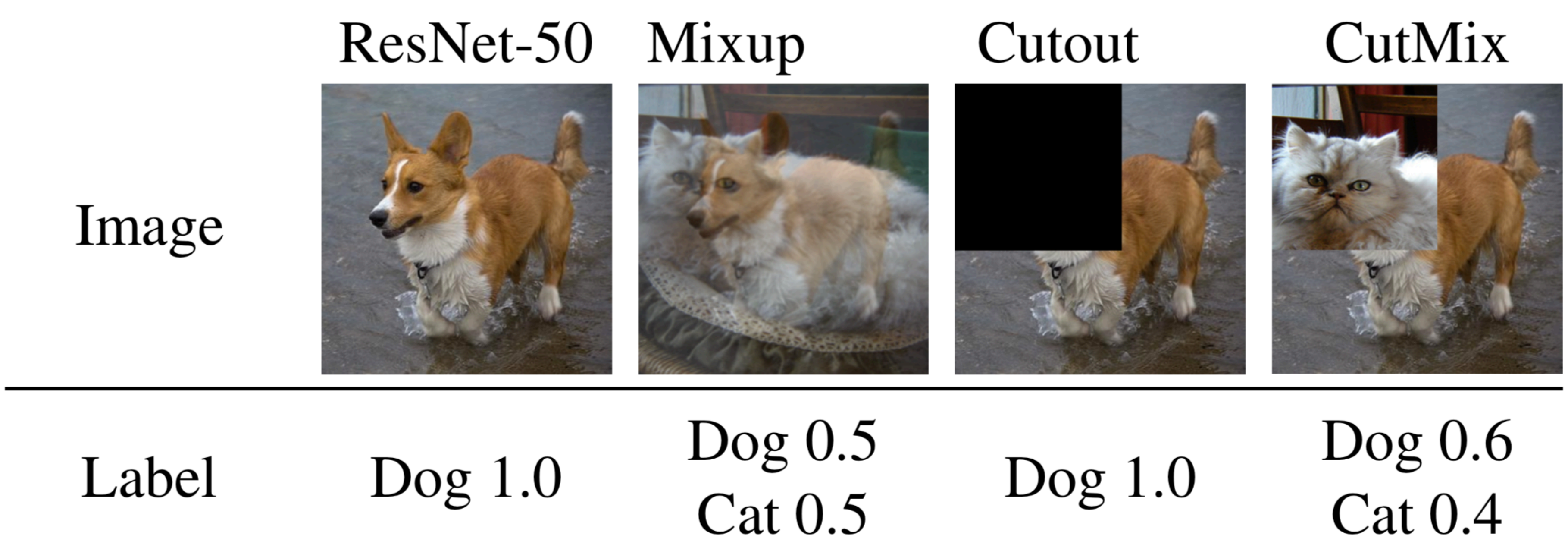
Sanghyuk Chun
Clova AI Research, Naver Corp.

Junsuk Choe
Yonsei University*

Youngjoon Yoo
Clova AI Research, Naver Corp.

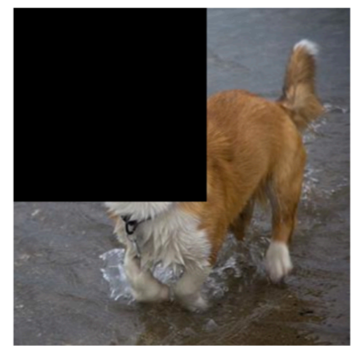
*** Visit researcher at Clova AI at the time.**

CutMix in a nutshell.



- Unlike Cutout, CutMix uses all input pixels for training.
- Unlike Mixup, CutMix presents realistic local image patches.
- Only 20 lines of code: https://github.com/ClovaAI/CutMix-PyTorch

Occlusion robustness and Positive side-effects.

	ResNet-50	Mixup	Cutout	CutMix
Image				
Label	Dog 1.0	Dog 0.5 Cat 0.5	Dog 1.0	Dog 0.6 Cat 0.4
ImageNet Cls (%)	76.3 (+0.0)	77.4 (+1.1)	77.1 (+0.8)	**78.4 (+2.1)**
ImageNet Loc (%)	46.3 (+0.0)	45.8 (-0.5)	46.7 (+0.4)	**47.3 (+1.0)**
Pascal VOC Det (mAP)	75.6 (+0.0)	73.9 (-1.7)	75.1 (-0.5)	**76.7 (+1.1)**

- Occ. Robustness
- Strong classifier
- Localizable feat.
- Pre-train model
    - Detection
    - Captioning

Classification performance.

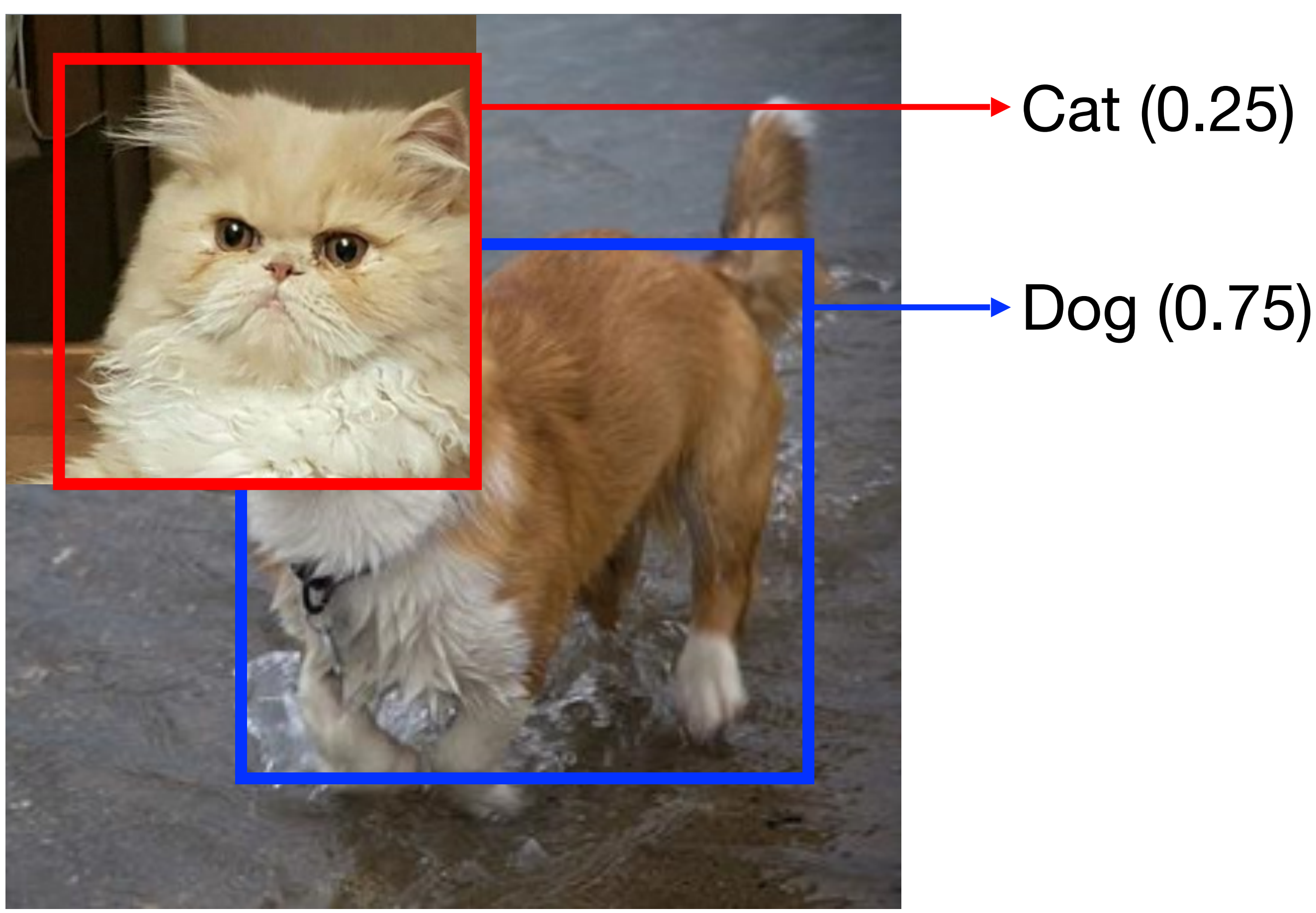
Model	# Params	Top-1 Err (%)	Top-5 Err (%)
ResNet-152*	60.3 M	21.69	5.94
ResNet-101 + SE Layer*	49.4 M	20.94	5.50
ResNet-101 + GE Layer*	58.4 M	20.74	5.29
ResNet-50 + SE Layer*	28.1 M	22.12	5.99
ResNet-50 + GE Layer*	33.7 M	21.88	5.80
ResNet-50 (Baseline)	25.6 M	23.68	7.05
ResNet-50 + Cutout	25.6 M	22.93	6.66
ResNet-50 + StochDepth	25.6 M	22.46	6.27
ResNet-50 + Mixup	25.6 M	22.58	6.40
ResNet-50 + Manifold Mixup	25.6 M	22.50	6.21
ResNet-50 + DropBlock*	25.6 M	21.87	5.98
ResNet-50 + Feature CutMix	25.6 M	21.80	6.06
ResNet-50 + CutMix	25.6 M	**21.60**	**5.90**
ResNet-50 + AutoAugment	25.6M	22.4*	6.2*

*** reported values from the reference paper**

- Great improvement over baseline.
- Better than existing regularizations.
- ResNet-50 + CutMix is better than ResNet-150.

Localizable Features.

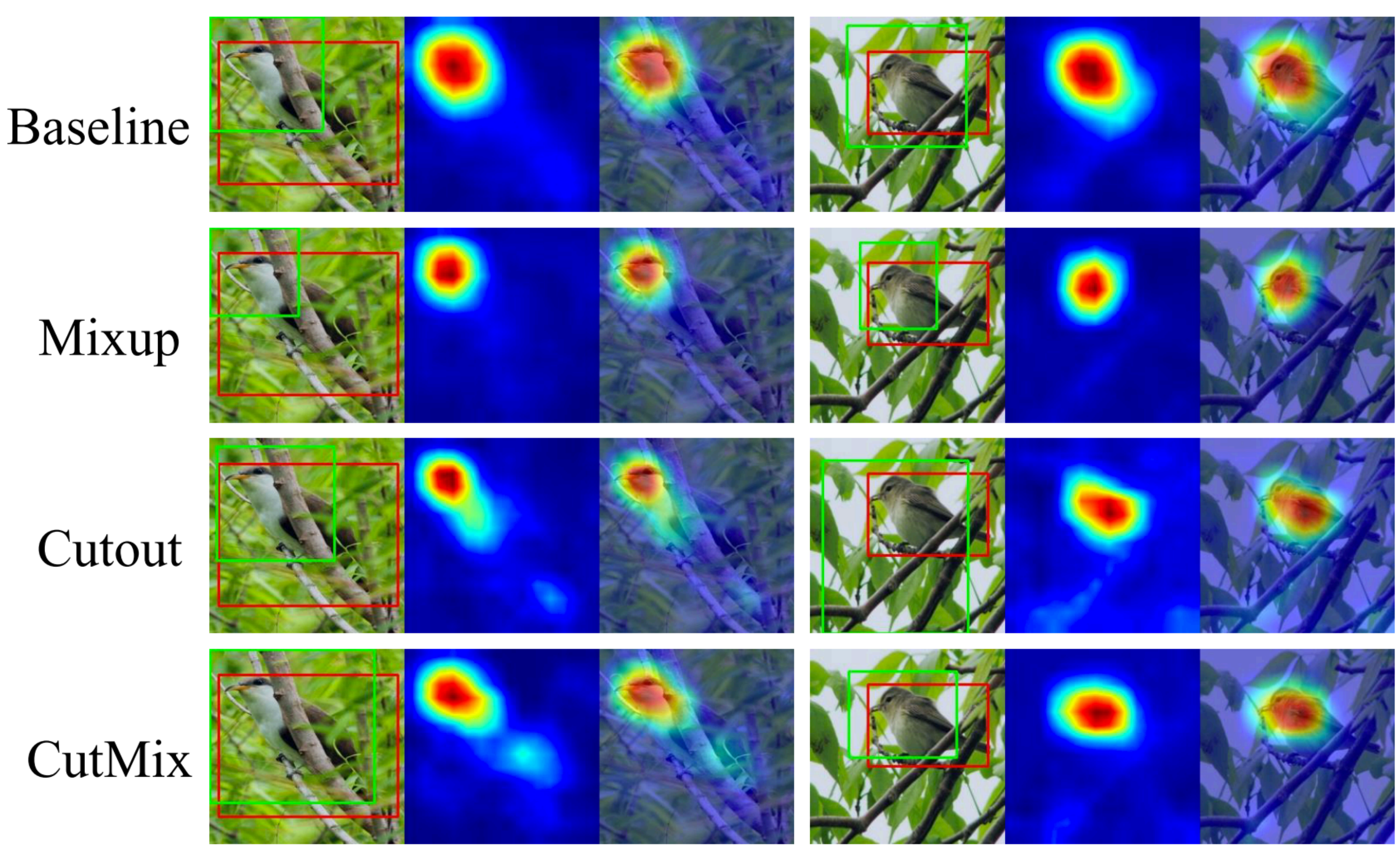
Method	CUB200-2011 Loc Acc (%)	ImageNet Loc Acc (%)
ResNet-50	49.41	46.30
Mixup	49.30	45.84
Cutout	52.78	46.69
CutMix	**54.81**	**47.25**



- CutMix makes model attend more "local" features unlike Mixup and Cutout.
- CutMix does not waste pixels during training.
- Great improvements in localization tasks

Localizable Features.

Method	CUB200-2011 Loc Acc (%)	ImageNet Loc Acc (%)
ResNet-50	49.41	46.30
Mixup	49.30	45.84
Cutout	52.78	46.69
CutMix	**54.81**	**47.25**



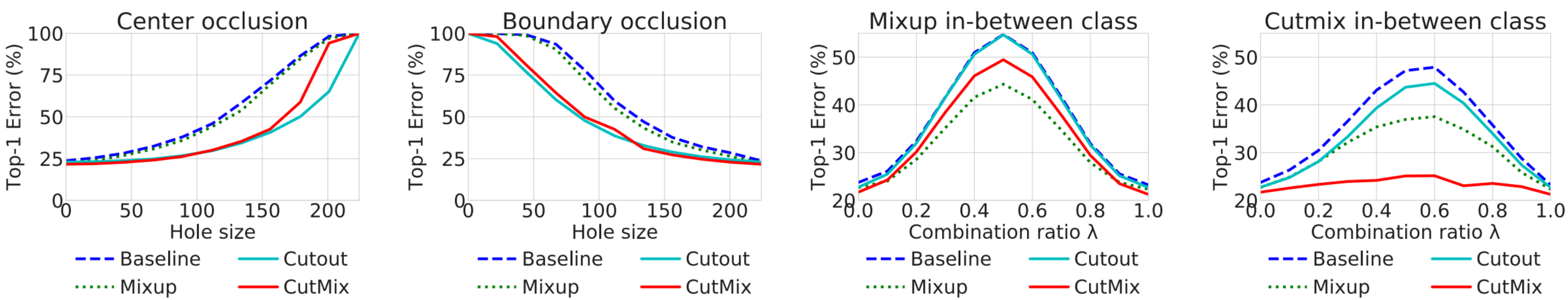
- CutMix makes model attend more "local" features unlike Mixup and Cutout.
- CutMix does not waste pixels during training.
- Great improvements in localization tasks

Transfer Learning.

Backbone Network	ImageNet Cls Top-1 Error (%)	Detection		Image Captioning	
		SSD [23] (mAP)	Faster-RCNN [29] (mAP)	NIC [41] (BLEU-1)	NIC [41] (BLEU-4)
ResNet-50 (Baseline)	23.68	76.7 (+0.0)	75.6 (+0.0)	61.4 (+0.0)	22.9 (+0.0)
Mixup-trained	22.58	76.6 (-0.1)	73.9 (-1.7)	61.6 (+0.2)	23.2 (+0.3)
Cutout-trained	22.93	76.8 (+0.1)	75.0 (-0.6)	63.0 (+1.6)	24.0 (+1.1)
CutMix-trained	21.60	**77.6 (+0.9)**	**76.7 (+1.1)**	**64.2 (+2.8)**	**24.9 (+2.0)**

- Localizability makes CutMix models attractive choices as pre-trained models.
- Improves tasks with localization elements: detection & captioning.

Robustness.



(a) Analysis for occluded samples

(b) Analysis for in-between class samples

	Baseline	Mixup	Cutout	CutMix
Top-1 Acc (%)	8.2	24.4	11.5	**31.0**

- CutMix shows better robustness than Mixup and Cutout in occlusion, in-between class samples and FGSM attack

Conclusion

- CutMix is a simple yet effective regularization technique for classification task
- CutMix shows better localization ability than previous methods such as Cutout, Mixup
- We observed that CutMix is effective for transfer learning, i.e., pre-training model for detection and captioning
- CutMix shows better robustness against occlusion, in-between class samples and adversarial noise

More details are in our paper!

- **CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, Youngjoon Yoo**

- https://arxiv.org/abs/1905.04899

- https://github.com/ClovaAI/CutMix-PyTorch

An Empirical Evaluation on Robustness and Uncertainty of Regularization Methods

Sanghyuk Chun
Clova AI Research,
Naver Corp.

Seong Joon Oh
Clova AI Research,
LINE Plus Corp.

Sangdoo Yun
Clova AI Research,
Naver Corp.

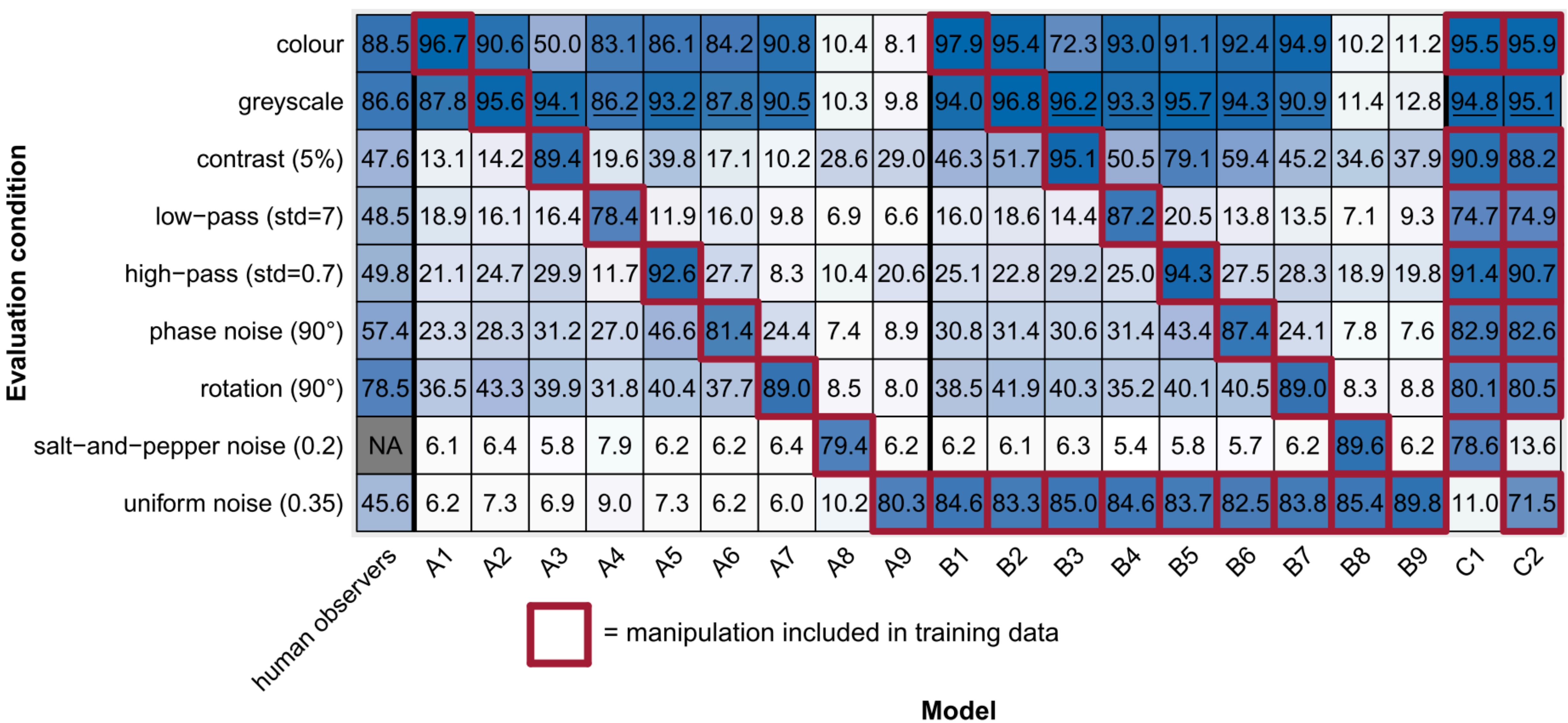
Dongyoon Han
Clova AI Research,
Naver Corp.

Junsuk Choe
Yonsei University*

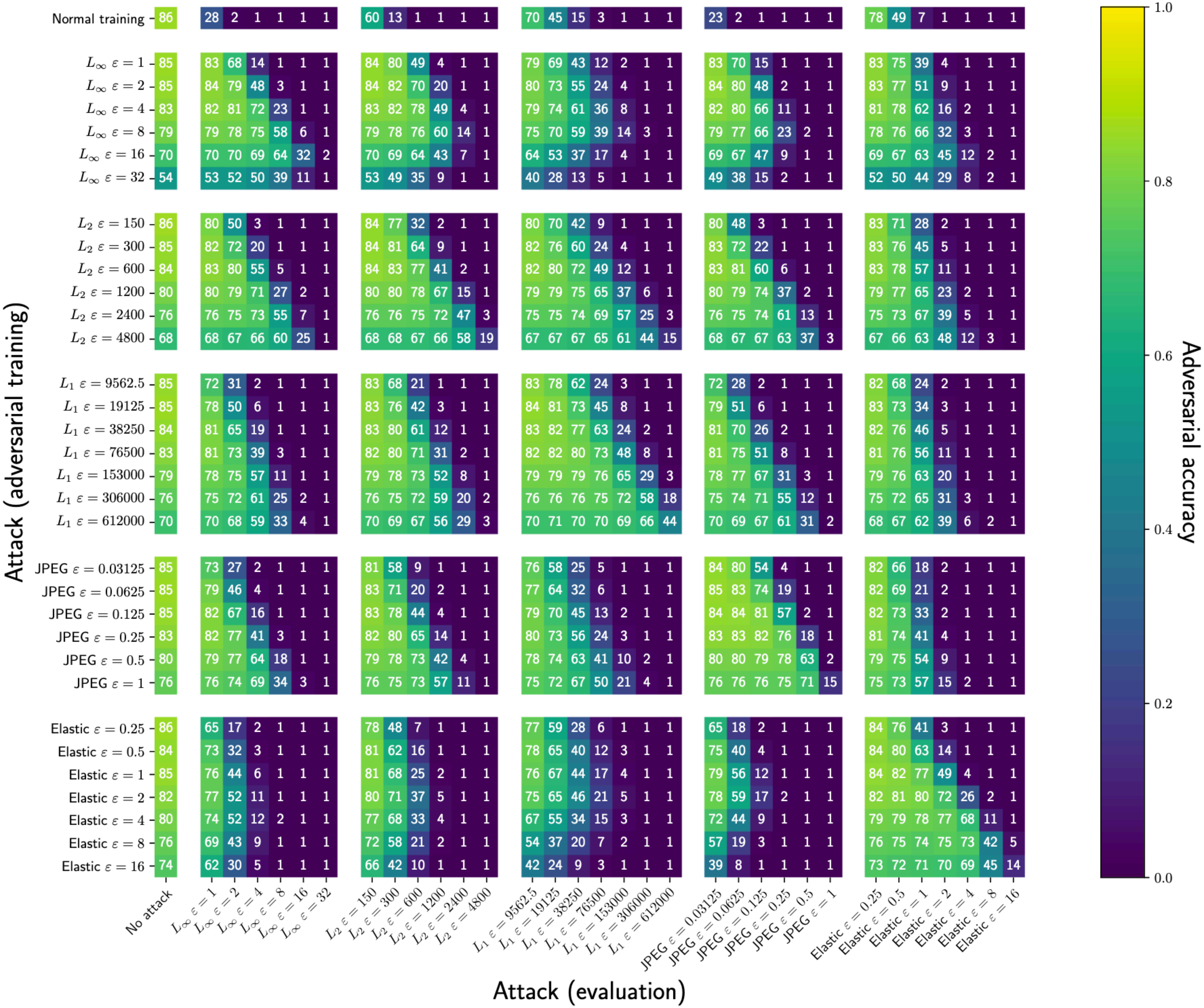
Youngjoon Yoo
Clova AI Research,
Naver Corp.

*** Visit researcher at Clova AI at the time.**

Generalization is matter.

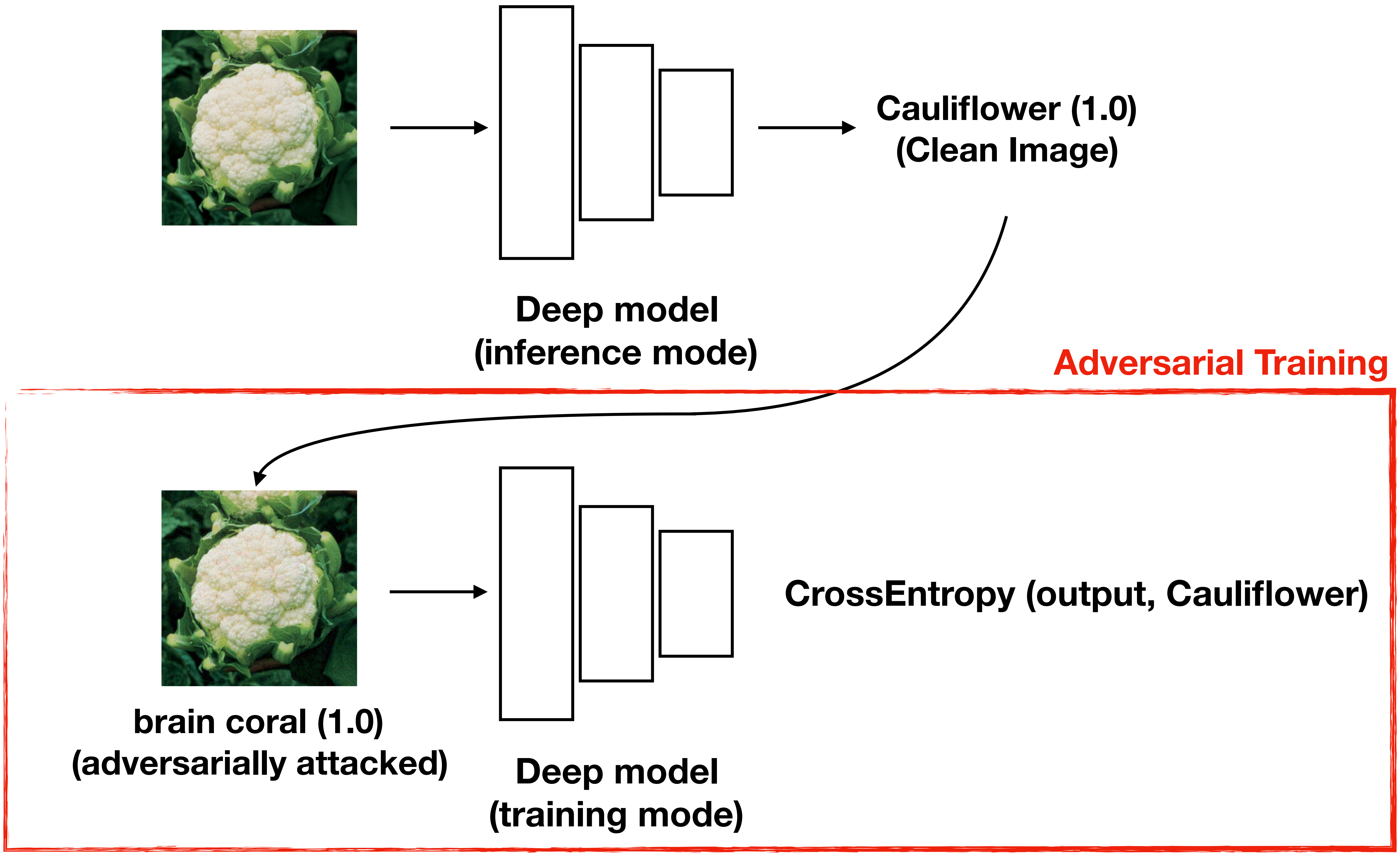


Geirhos, Robert, et al. "Generalisation in humans and deep neural networks." Advances in Neural Information Processing Systems. 2018.



Kang, Daniel, et al. "Transfer of Adversarial Robustness Between Perturbation Types." ICML 2019 UDL workshop

Current solutions are complicated and expensive: Adversarial training.



Current solutions are complicated and expensive: Adversarial training.

- Improve robustness by solving expensive minimax problem

$$\min_{\theta} \sum_{i}^{n} \max_{\varepsilon \in \mathcal{E}} \ell(f_\theta(x+\varepsilon, y))$$

Inner max problem is generally approximated by adversarial attacks: They are too expensive at scale

Training Speed:

With 30 attack iterations during training, the `Res152 Baseline` model takes about 52 hours to finish training on 128 V100s.

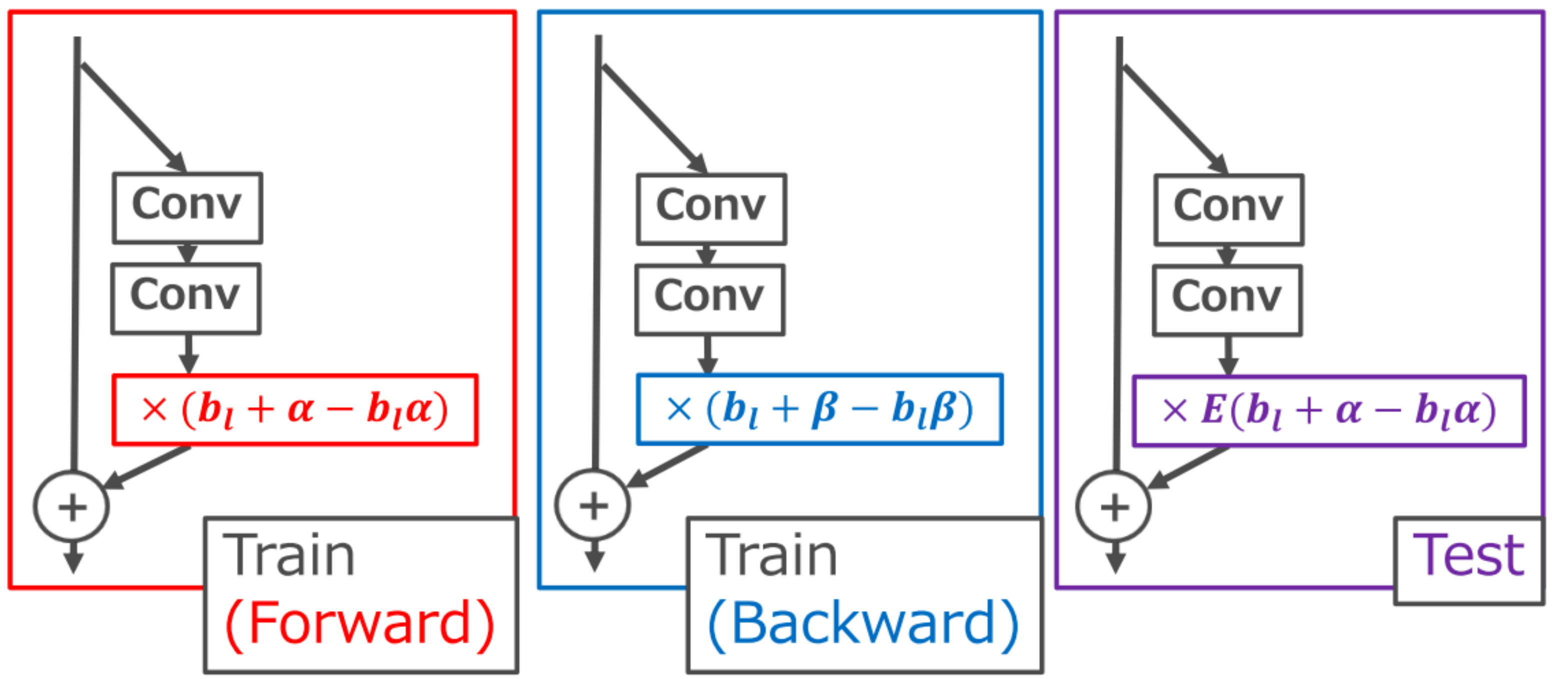
Under the same setting, the `Res152 Denoise` model takes about 90 hours on 128 V100s. Note that the model actually does not add much computation to the baseline, but it lacks efficient GPU implementation for the softmax version of non-local operation. The dot-product version, on the other hand, is much faster.

https://github.com/facebookresearch/ImageNet-Adversarial-Training/blob/master/INSTRUCTIONS.md

There are many cheap and effective regularization methods

- Augmentation methods:
  - Cutout, Mixup, CutMix
- Randomly feature drop:
  - Dropout, DropBlock, ShakeShake, ShakeDrop
- Label noise
  - Label smoothing, Mixup, CutMix
- In this talk, we do not consider the methods with additional parameters such as SE block, GE block

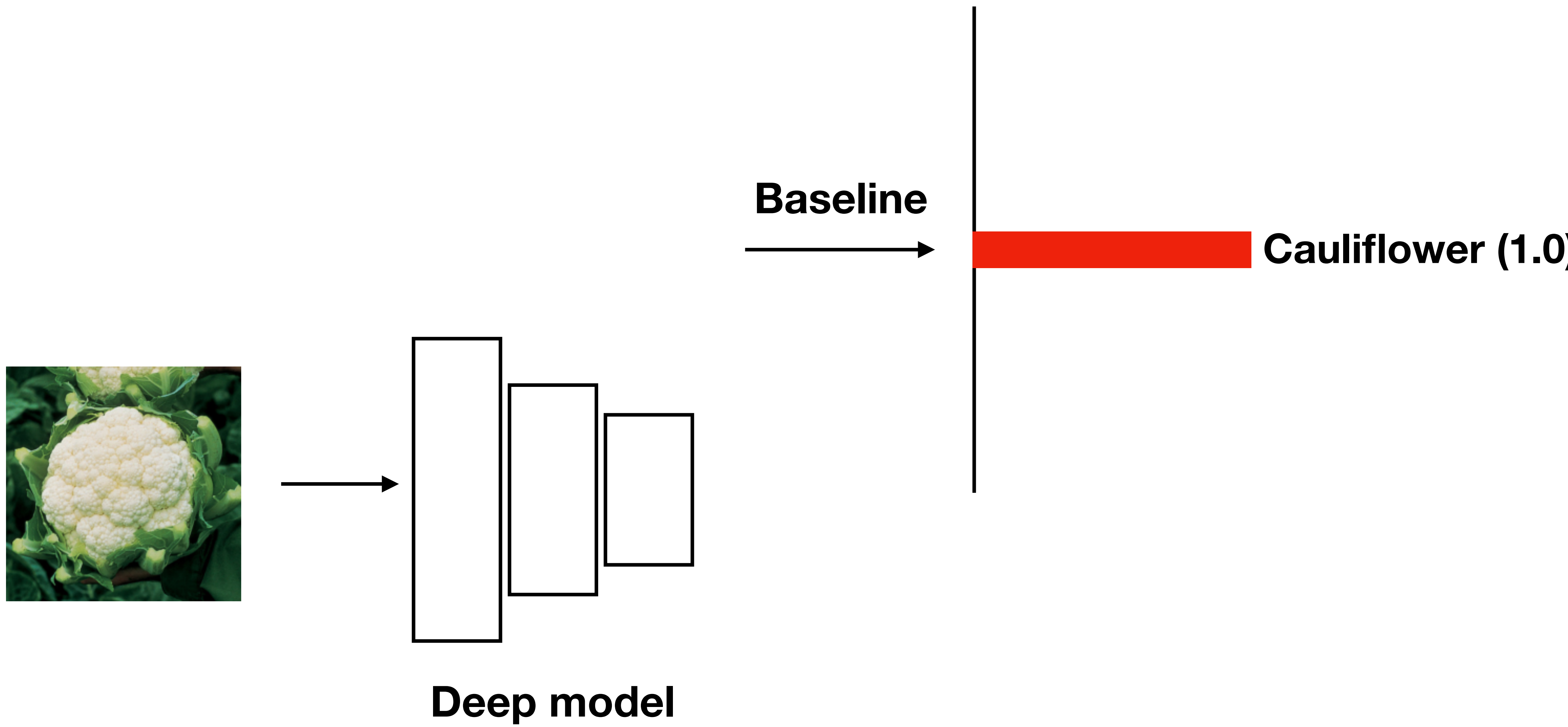
Selected regularization methods: ShakeDrop



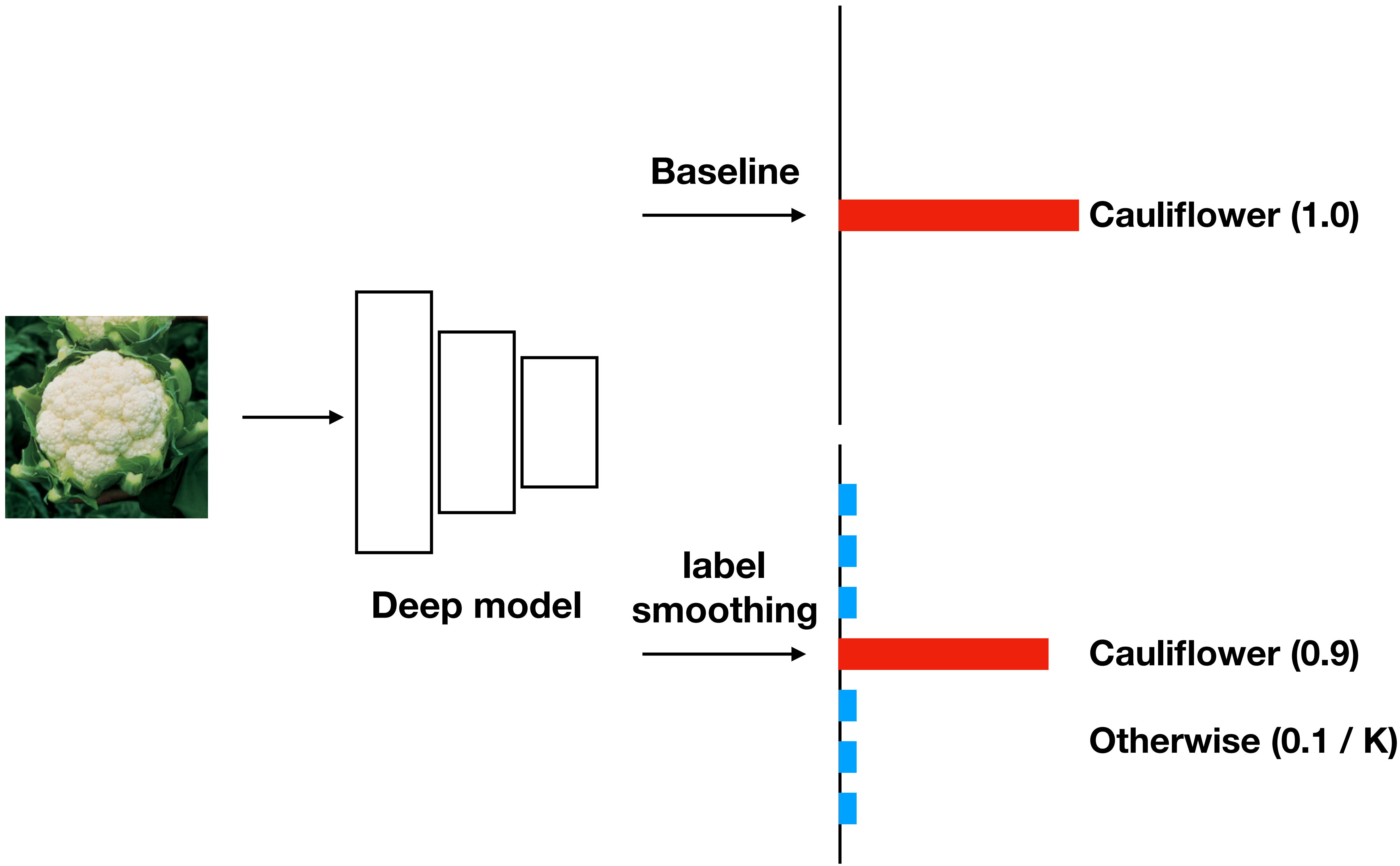
(d) ShakeDrop for 2- and 3-branch ResNet family

$$G(x) = \begin{cases} x + \alpha F_1(x) + (1-\alpha)F_2(x), & \text{in train-fwd} \\ x + \beta F_1(x) + (1-\beta)F_2(x), & \text{in train-bwd} \\ x + E[\alpha]F_1(x) + E[1-\alpha]F_2(x), & \text{in test,} \end{cases}$$

Selected regularization methods: Label smoothing



Selected regularization methods: Label smoothing



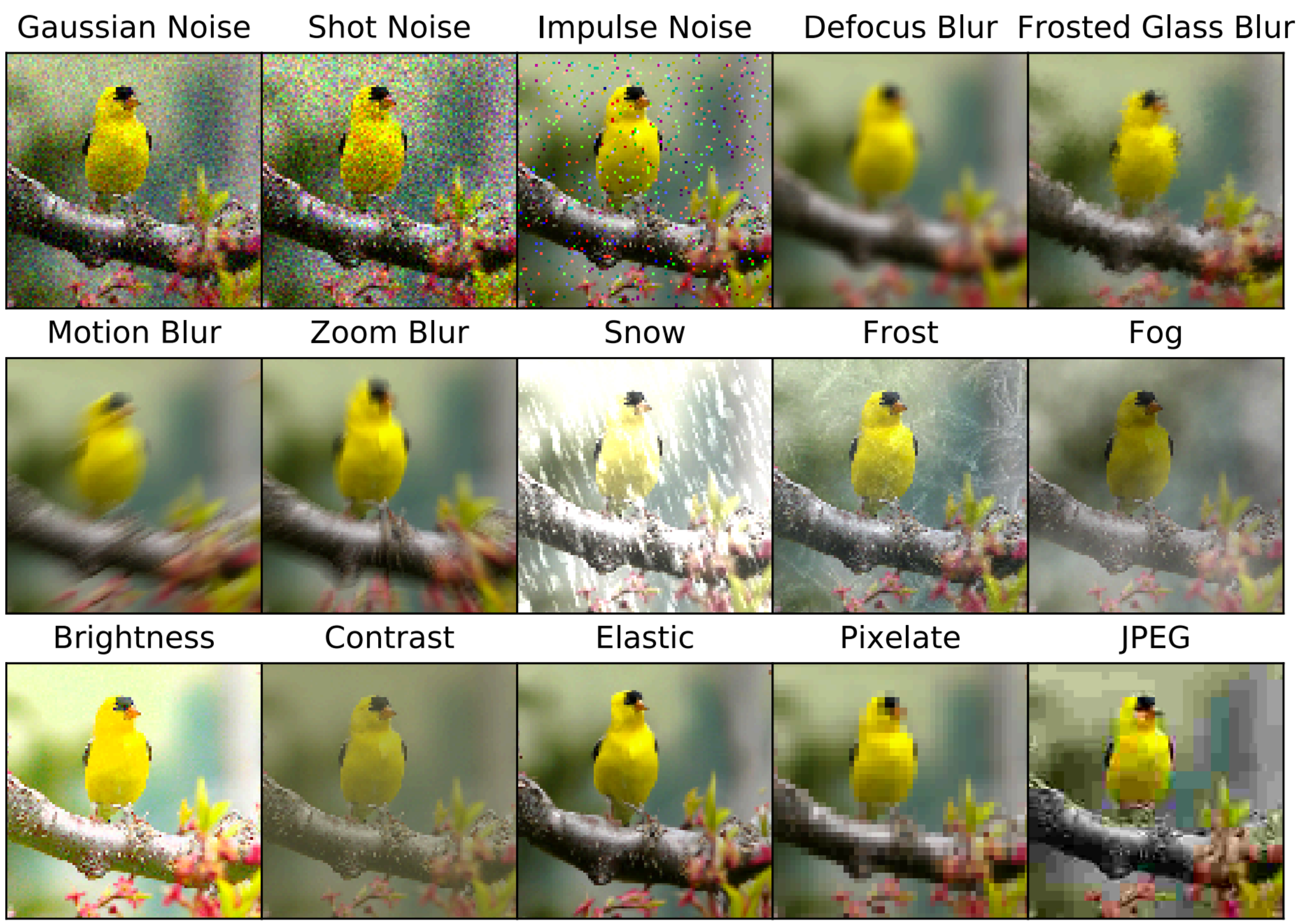
Benchmark 1: Adversarial robustness

- FGSM (Fast Gradient Sign Method)

- Note: regularization methods can not provide provable defense to adversarial robustness

Benchmark 2: Non-adversarial robustness

- Occlusion
- ImageNet-C: Noise, blur, weather change, digital



Hendrycks, Dan, and Thomas Dietterich. "Benchmarking neural network robustness to common corruptions and perturbations." ICLR 2019

CIFAR-100 Results

Methods	CIFAR-100 Top-1 Err.	FGSM Top-1 Err.	Occlusion Top-1 Err.	CIFAR-C mCE	Noise Top-1 Err.	Blur Top-1 Err.	Weather Top-1 Err.	Digital Top-1 Err.
Baseline (PyramidNet-200)	**16.45**	84.20	72.19	45.11	74.62	**46.77**	**30.66**	38.65
Adversarial Logit Pairing	24.75	**51.32**	92.27	50.04	69.94	51.75	40.62	44.70
Cutout	16.53	91.07	**27.00**	51.65	89.77	51.40	34.24	43.20
Add Gaussian Noise	19.49	85.08	73.23	**42.01**	**54.63**	48.42	31.54	**38.48**

- Observation: a targeted solution only solves the targeted problem (e.g., Cutout only improves occlusion robustness while worsen FGSM and CIFAR-100-C robustness)
- A similar result is shown by Geirhos, et al., 2018

Geirhos, Robert, et al. "Generalisation in humans and deep neural networks." Advances in Neural Information Processing Systems. 2018.

CIFAR-100 Results

Method	CIFAR-100 Top-1 Err.	FGSM Top-1 Err.	CIFAR-C Top-1 Err.	Occlusion Top-1 Err.
Baseline (PyramidNet-200)	16.45	84.20	45.11	72.19
Cutout + SD + LS	13.49	69.59	43.86	26.33
Mixup + SD + LS	14.79	56.32	40.32	56.76
CutMix + SD + LS	13.83	62.72	44.99	34.96
Adversarial Logit Pairing	24.75	51.32	50.04	92.27
Add Gaussian Noise	19.49	85.08	42.01	73.23
OOD augment (SVHN)	38.80	97.35	67.03	79.13
OOD augment (GAN)	34.78	94.65	57.09	85.30

- Good regularization methods are strong baselines, i.e., they are "generally" better than specific solutions and the baseline.

ImageNet Results

	Average	Clean	FGSM	Occ.	Noise	Blur	Weather	Digital	mCE
Baseline (ResNet-50)	67.43	23.68	91.85	46.01	78.58	86.63	64.99	80.24	77.55
Label Smoothing	62.67	22.31	73.60	44.35	77.08	82.30	61.72	77.33	74.44
ShakeDrop	64.57	22.03	87.19	42.98	76.13	83.42	61.56	78.69	74.87
ShakeDrop + LS	61.45	21.92	72.65	42.85	74.47	82.15	60.47	75.67	73.10
Cutout	64.81	22.93	88.50	29.72	79.94	85.37	65.34	81.87	78.01
Cutout + LS	61.90	22.02	75.24	29.08	79.80	84.51	62.72	79.93	76.54
Mixup	61.46	22.58	75.60	44.20	73.09	81.49	58.83	74.42	71.88
Mixup + LS	58.54	22.41	69.43	42.31	65.36	82.95	53.37	73.94	69.14
CutMix	62.08	21.60	69.04	30.09	80.88	84.87	64.11	83.95	78.29
CutMix + LS	61.02	21.87	67.41	31.51	77.01	84.61	63.13	81.56	76.55
CutMix + SD	61.75	21.60	80.00	31.28	77.06	84.18	61.04	77.07	74.69
CutMix + SD + LS	60.96	21.90	68.65	31.62	76.04	84.53	62.82	81.16	76.14

- Largely similar to CIFAR-100 results
- We observe that Mixup + LS shows the best performance in ImageNet-C mCE than other expensive methods

Conclusion

- Simple regularization techniques are effective in enhancing robustness and uncertainty estimation.
- Well-regularized models achieve state-of-the-art robustness (e.g., 69.14% mCE for Mixup + LS).
- Methods for specific tasks (e.g., adversarial training, Cutout) do not generalize to other tasks.
- State-of-the-art regularization methods (e.g., Cutout, Mixup, CutMix, ShakeDrop, label smoothing) should be considered as powerful baselines.

More details are in our paper!

- **An Empirical Evaluation on Robustness and Uncertainty of Regularization Methods. Sanghyuk Chun, Seong Joon Oh, Sangdoo Yun, Dongyoon Han, Junsuk Choe, Youngjoon Yoo**

- Presented in ICML 2019 Uncertainty & Robustness in Deep Learning Workshop (Friday)

Side Topic:
Robustness in
non-vision data (music).

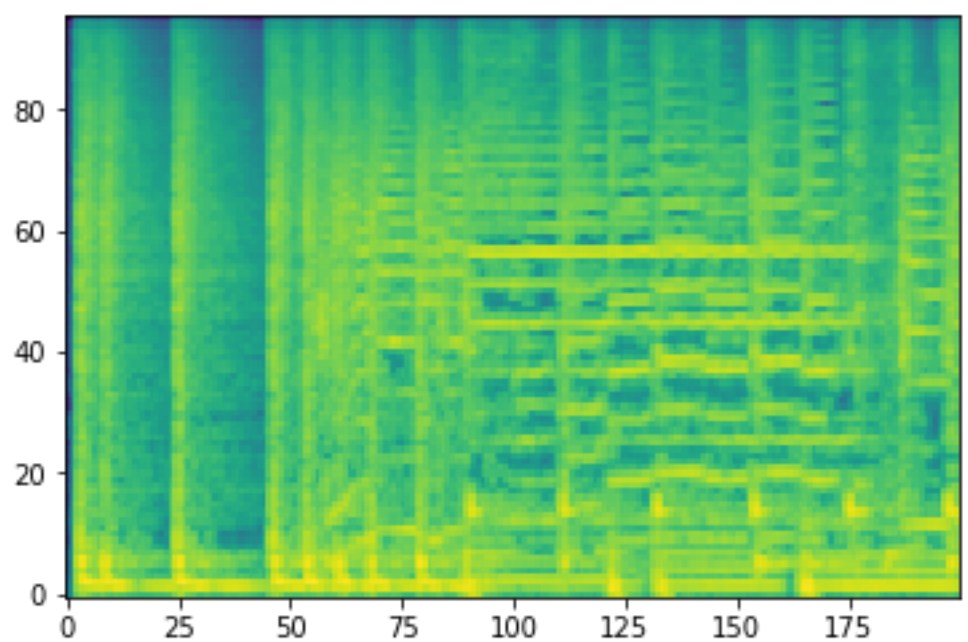
Visualizing and Understanding Self-attention based Music Tagging

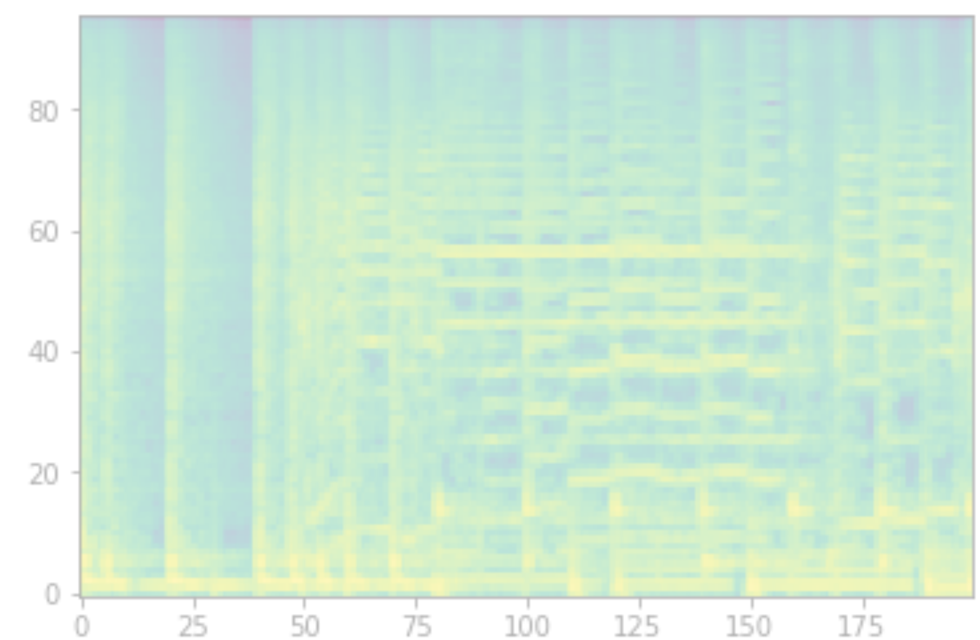
Minz Won
Music Technology Group,
Universitat Pompeu Fabra

Sanghyuk Chun
Clova AI Research,
Naver Corp.

Xavier Serra
Music Technology Group,
Universitat Pompeu Fabra

Also matters to other domains; Music Understanding.

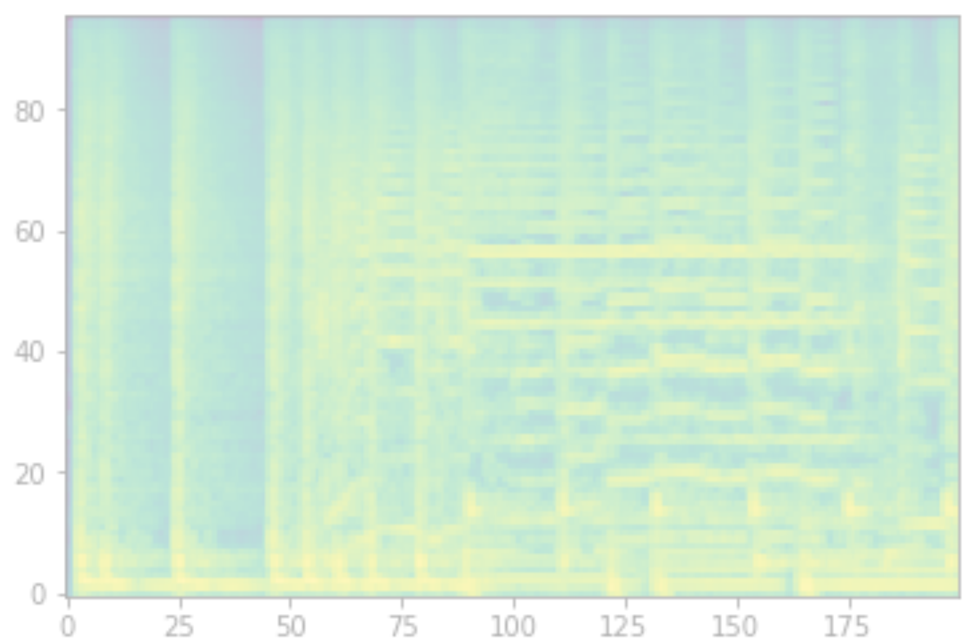




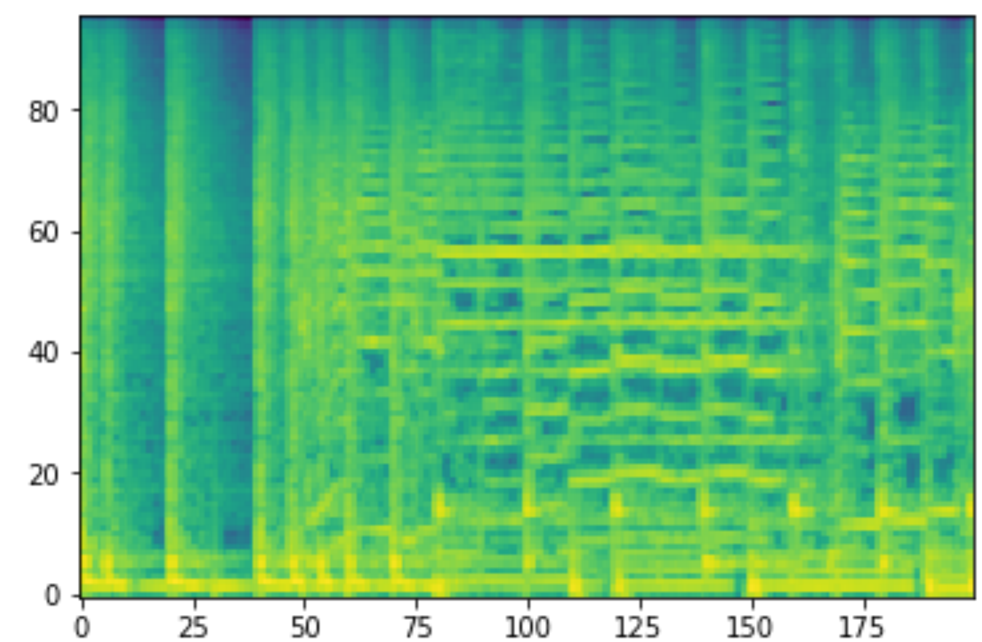
124 BPM
Predicted to ChaChaCha (correct)

130 BPM
Predicted to Tango (fooled)

Also matters to other domains; Music Understanding.



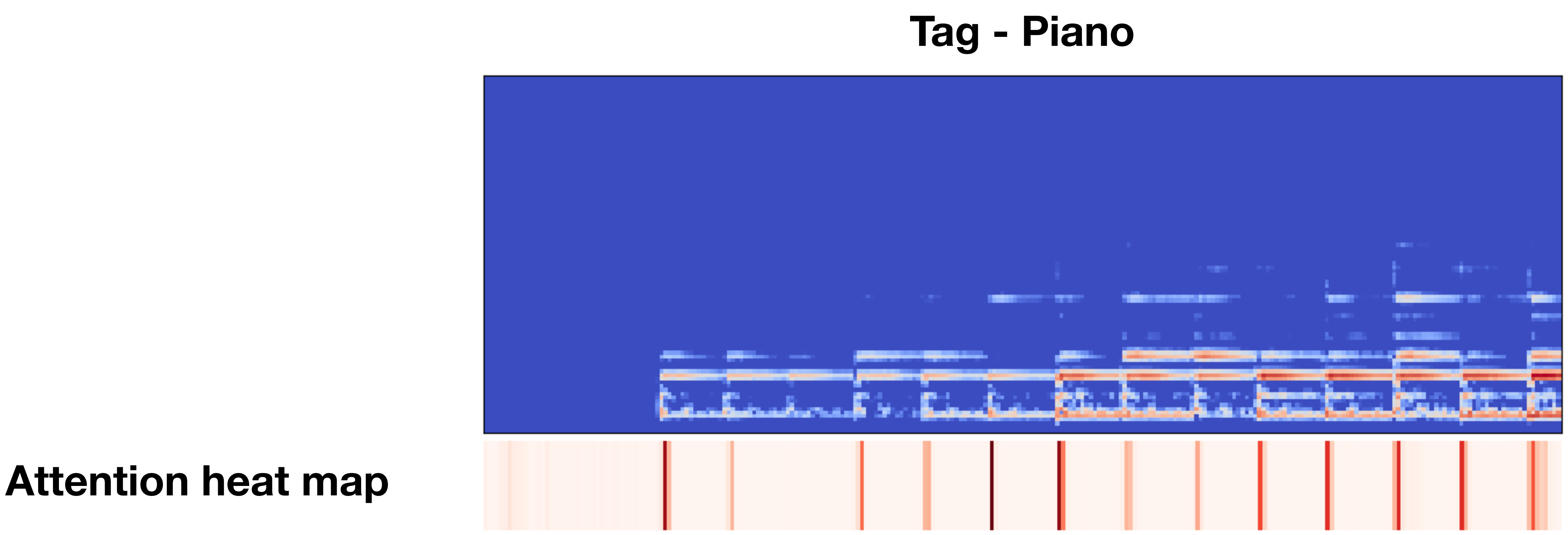
124 BPM
Predicted to ChaChaCha (correct)



130 BPM
Predicted to Tango (fooled)

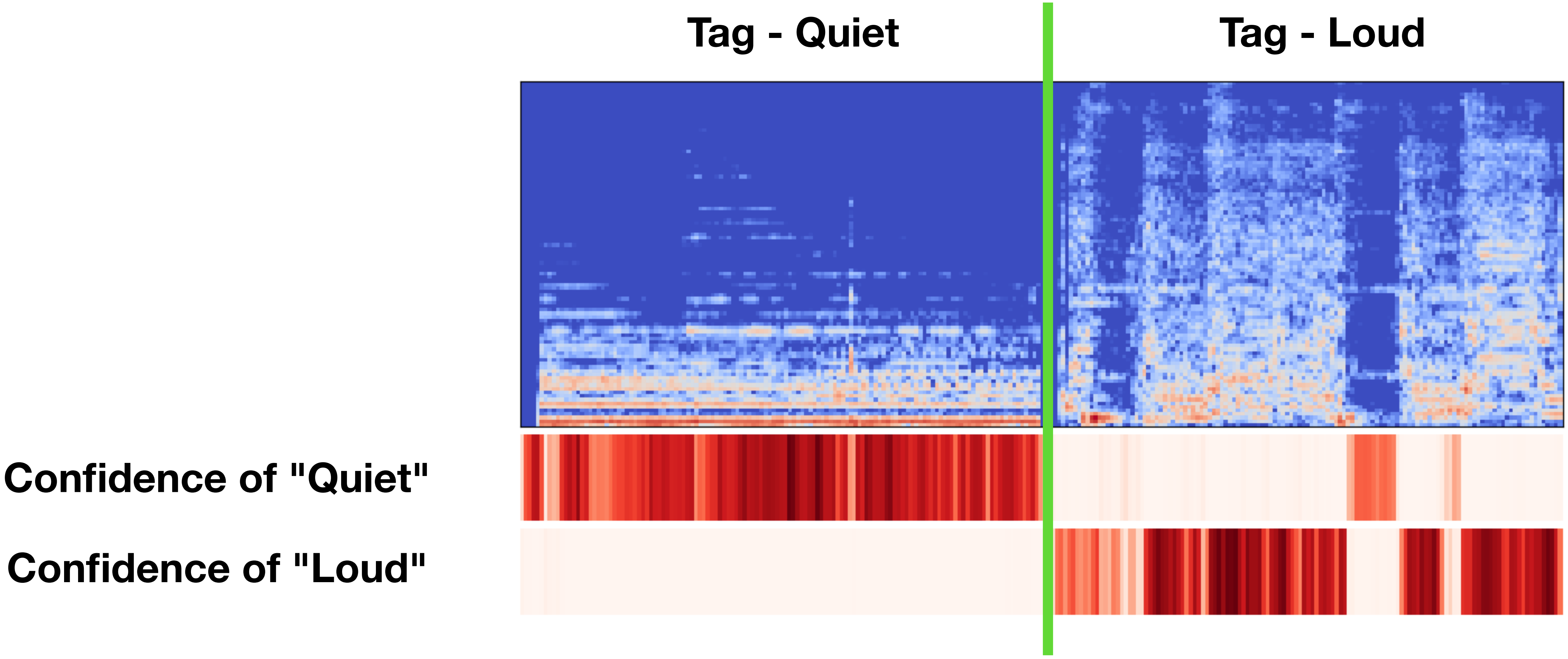
Interpretability is the matter; Where is attended by the model?

- Observation 1: Model focuses on "energy"



Interpretability is the matter; Where is attended by the model?

- Observation 2: Models understand the music with only small chunks



More details are in our paper!

- **Visualizing and Understanding Self-attention based Music Tagging. Minz Won, Sanghyuk Chun, Xavier Serra**

- Presented in ICML 2019 Machine Learning for Music Discovery Workshop (Contribute Talk, Saturday)

Conclusion and future works.

Conclusion and future works.

- Training strategy changes the property of models **without any changes in architectures**
  - e.g., adversarial training, CutMix, ...
- The direct noise augmentation is a good solution to the specific robustness problem but it **cannot be generalized.**
- We should consider not only specific robustness but also **the generalization ability of deep models** for future works.

See you at...

NAVER & LINE Booth #111 (SUN, MON, TUE, WED)

Poster and Oral talk for "Curiosity-Bottleneck: Exploration By Distilling Task-Specific Novelty" (TUE)

Poster session at UDL workshop, "An Empirical Evaluation on Robustness and Uncertainty of Regularization Methods" (FRI)

Contributed talk at ML4DL workshop, "Visualizing and Understanding Self-attention based Music Tagging" (SAT)

Internship & full-time opportunities at Clova.

- We do lots of exiting researches at Clova AI!

Machine Learning
- **Lightweight models**
- **Regularization methods**
- **Uncertainty estimation**
- **ML Robustness & adversarial learning**
- **AutoML**
- **Reinforcement learning**

Computer Vision
- **OCR**
- **Detection & segmentation (object, human, face)**
- **Pose estimation & action recognition**
- **Generative models**

Natural Language Processing
- **Large-scale language model**
- **Goal-oriented dialog**

Internship & full-time opportunities at Clova.

- Positions: Research Scientist / AI Software Engineer / Research Internship / Global Residency
- Job descriptions: https://clova.ai/en/research/careers.html
- Please contact via clova-jobs@navercorp.com