MLLAB 머신러닝 연구실

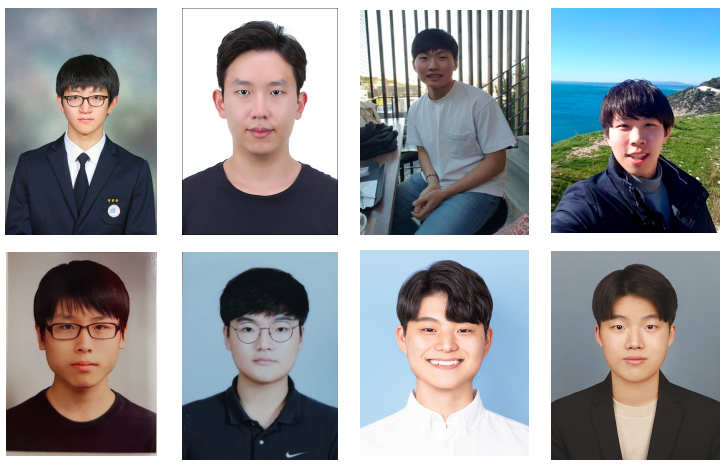
homepage

Faculty

송현오 교수님

- Associate Professor, SNU
- Research Scientist, Google Research
- Postdoc, Stanford University
- Ph.D. in Computer Science, UC Berkeley

Students



7 MS/Ph.D. students + 1 Research Intern

- 3 B.S. in Mathematical Science
- 2 B.S. in Statistics
- 1 B.S. in Electrical and Computer engineering
- 1 B.S. in Computer Science and Engineering
- 1 B.S. in Economics

Alumni

- Hyoungseok Kim M.S. (2018.09 ~ 2020.08)
  - Now at Unnoted
- Wonho Choo M.S. (2020.03 ~ 2022.02)
  - Now at Kakao

Publication

We publish papers on major machine learning conferences

- In 2022, 2 ICML, 1 NeurIPS, 1 AAAI, 1 AISTATS
- In 2021, 1 ICLR (oral), 1 NeurIPS
- In 2020, 1 ICML
- In 2019, 3 ICML (2 long talk), 1 CVPR
- In 2018, 1 ICML (1 long talk)

Environment

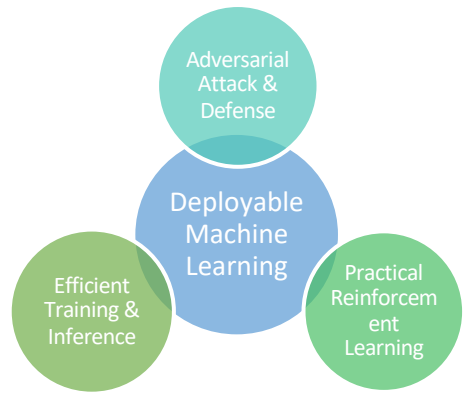
Lab office

- Samsung Electronics-SNU Research Center (7F, Building 944)

Lab server

- 50+ servers with 200+ GPUs
- Slurm job manager

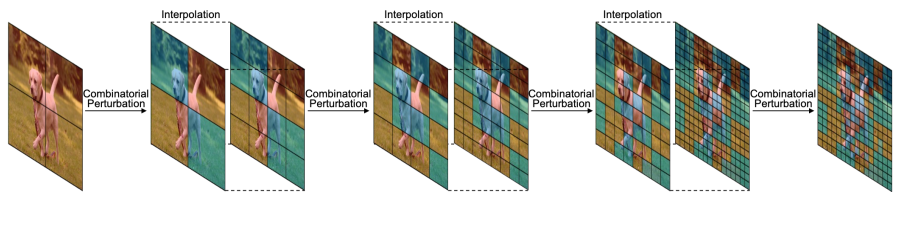
Research



Adversarial Attack & Defense

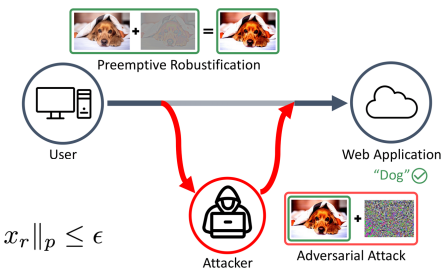
- Black-box adversarial attack
  on Image domain [ICML19]
  on language domain [ICML22]

$$\underset{x_{adv}}{\text{maximize}}\ f(x_{adv})$$
$$\text{subject to}\ x_{adv} - x \in \{\epsilon, -\epsilon\}^p$$



- Preemptive robustification of data [AAAI22]

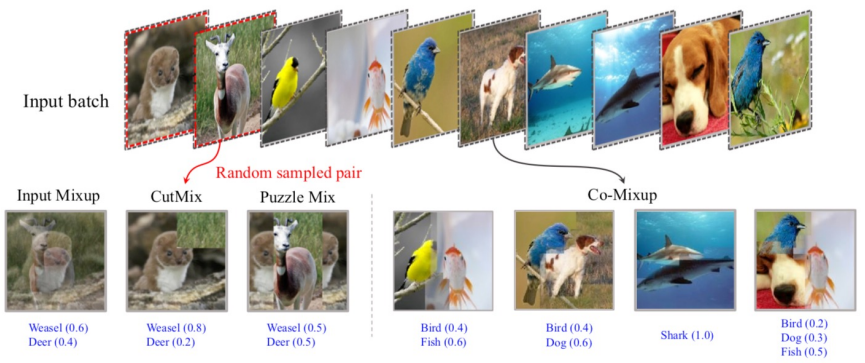
$$\underset{x_r}{\text{minimize}}\ \sup_{x_r^a}\ \ell(x_r^a, c(x_o))$$
$$\text{subject to}\ \|x_r - x_o\|_p \le \delta\ \text{and}\ \|x_r^a - x_r\|_p \le \epsilon$$



Efficient Training & Inference

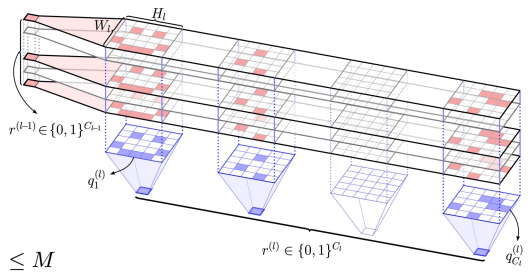
- Saliency-based data augmentation [ICML20] and its batch-level extension [ICLR21]

$$\underset{z_{j,k}\in\mathcal{L}^m,\ \|z_{j,k}\|_1=1}{\text{argmin}} \sum_{j=1}^{m'}\sum_{k=1}^{n} c_k^\intercal z_{j,k} + \beta \sum_{j=1}^{m'}\sum_{(k,k')\in\mathcal{N}} (1 - z_{j,k}^\intercal z_{j,k'}) + \gamma \max\left\{\tau,\ \sum_{j=1}^{m'}\sum_{j'\neq j}^{m'}\left(\sum_{k=1}^{n} z_{j,k}\right)^\intercal A \left(\sum_{k=1}^{n} z_{j',k}\right)\right\}$$



- Pruning neural networks [AISTAT22] and dataset condensation [ICML22]

$$\underset{r^{(0:L)}}{\text{maximize}}\ \sum_{l=1}^{L}\left\langle I^{(l)}, A^{(l)}\right\rangle$$
$$\text{subject to}\ \sum_{l=0}^{L} a_l \left\|r^{(l)}\right\|_1 + \sum_{l=1}^{L} b_l \left\|A^{(l)}\right\|_1 \le M$$
$$A^{(l)} = r^{(l-1)} r^{(l)\intercal} \otimes J_{K_l} \quad \forall l \in [L].$$

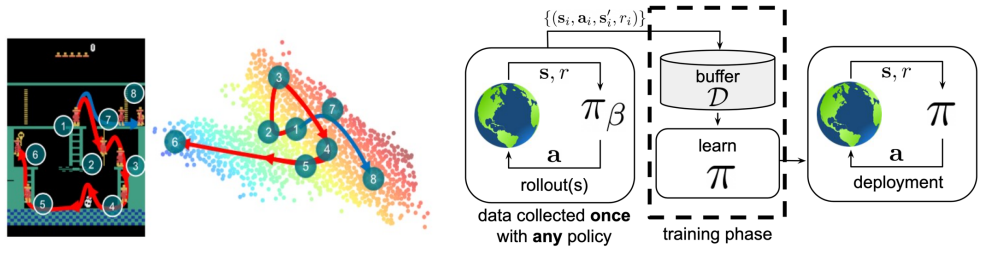


- Representation learning [ICML18, CVPR19] and disentanglement [ICML19]

Practical Reinforcement Learning

- RL exploration [ICML2019], Offline RL algorithm [NeurIPS2021], and RL generalization [NeurIPS2022]

$$\min_{\phi_i} \mathbb{E}_{\mathbf{s},\mathbf{a},\mathbf{s}'\sim\mathcal{D}}\left[\left(Q_{\phi_i}(\mathbf{s},\mathbf{a}) - \left(r(\mathbf{s},\mathbf{a}) + \gamma\, \mathbb{E}_{\mathbf{a}'\sim\pi_\theta(\cdot|\mathbf{s}')}\left[\min_{j=1,\ldots,N} Q_{\phi_j'}(\mathbf{s}',\mathbf{a}') - \beta \log \pi_\theta(\mathbf{a}'\mid\mathbf{s}')\right]\right)\right)^2\right]$$



2022.09