November 2024

TiDeL Hackathon

Presenter

Rahul Vigneswaran K

CS23MTECH02002

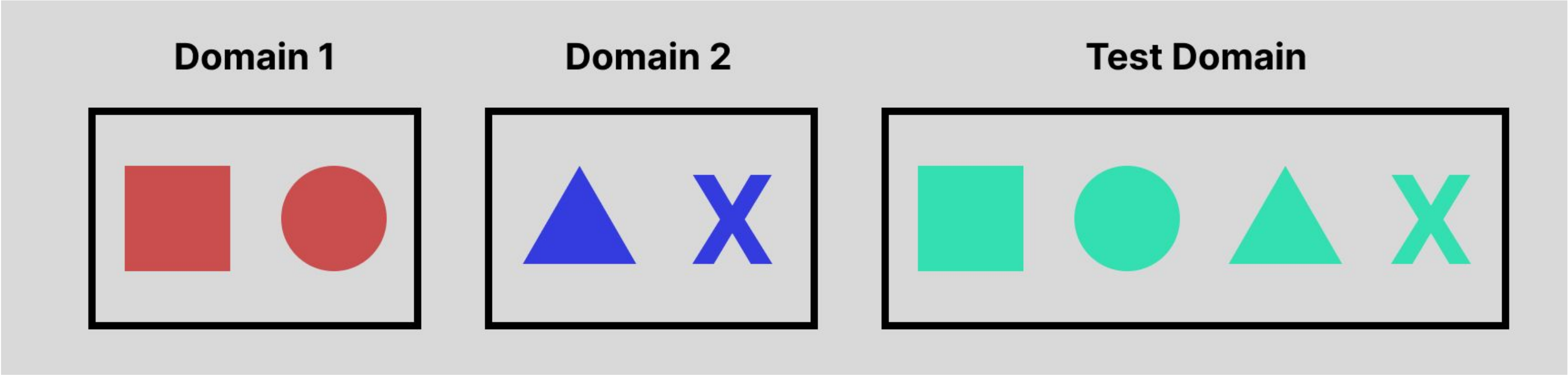
Course Instructor

Dr. Srijith PK

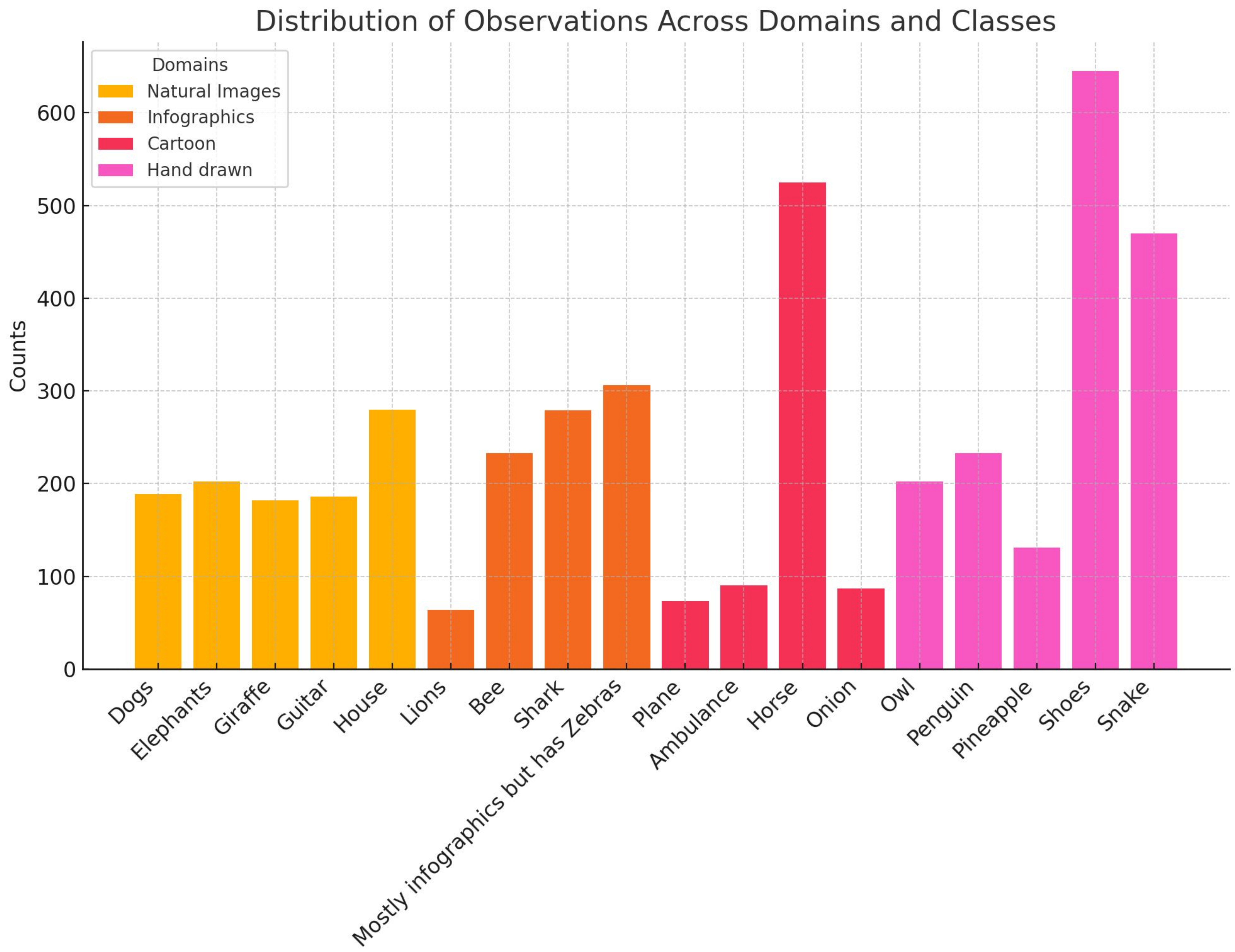
*Presented as part of the coursework for Topics in Deep Learning (CS6420)

Table Of Contents

Setup

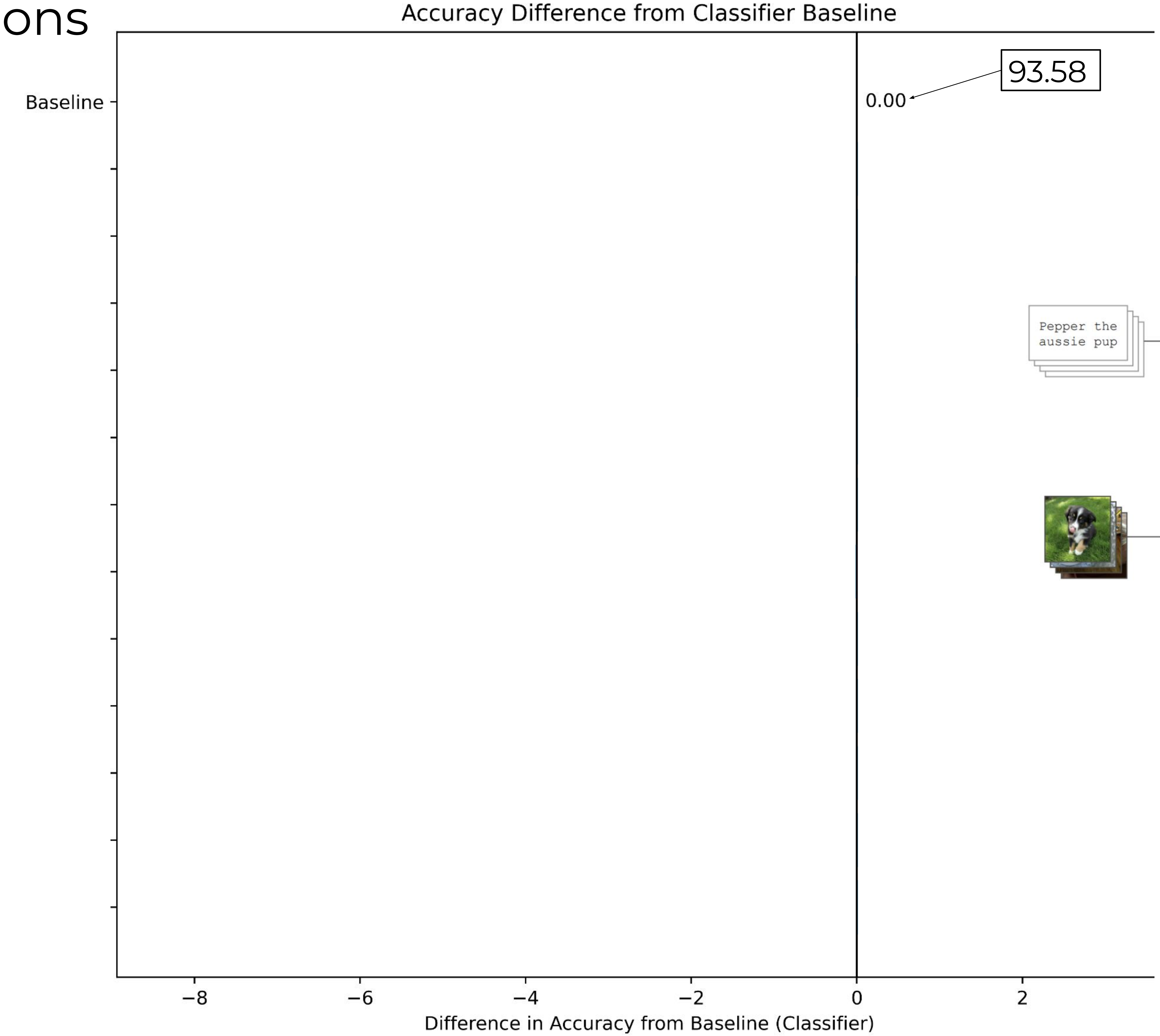


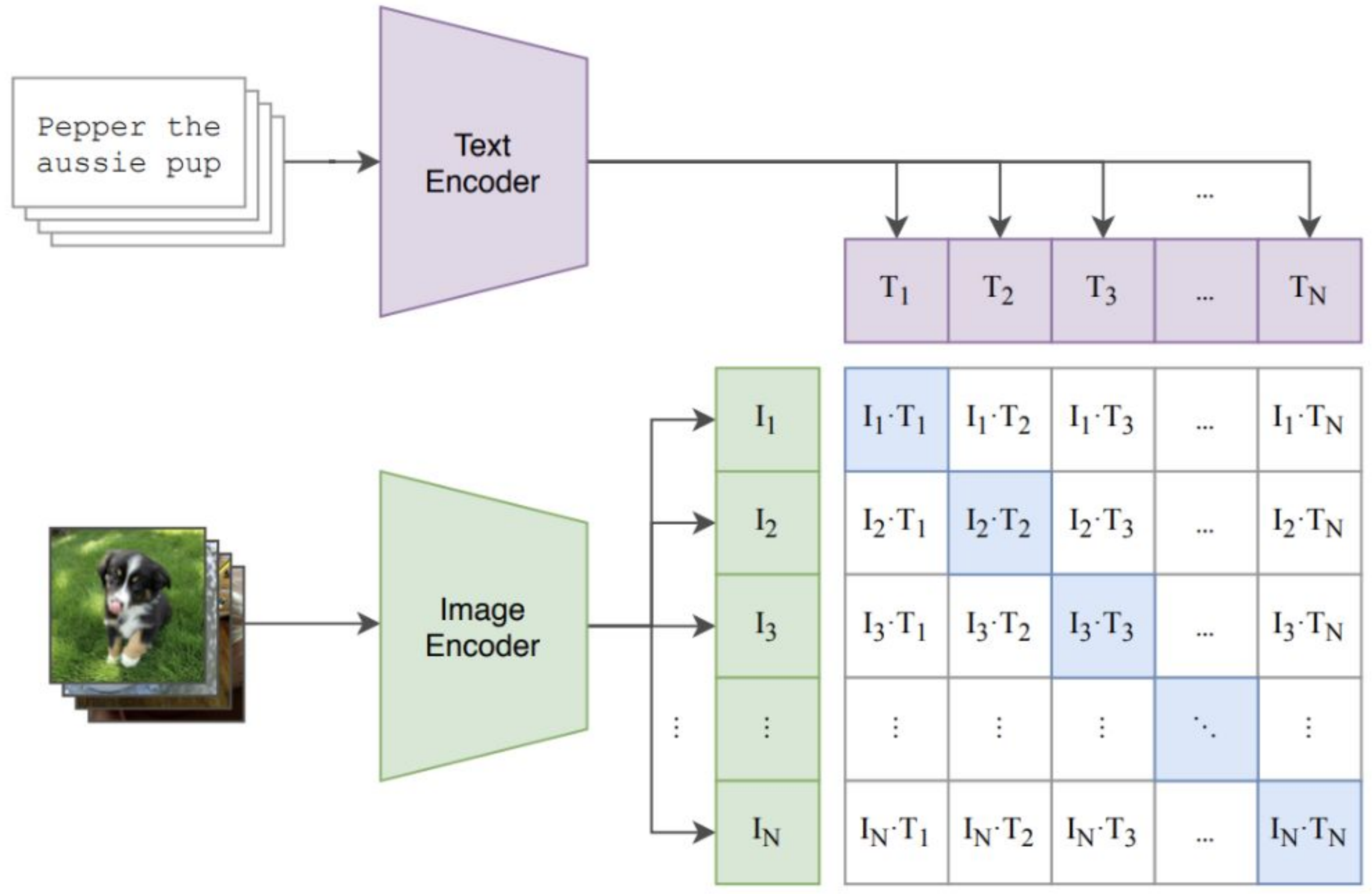
Observations



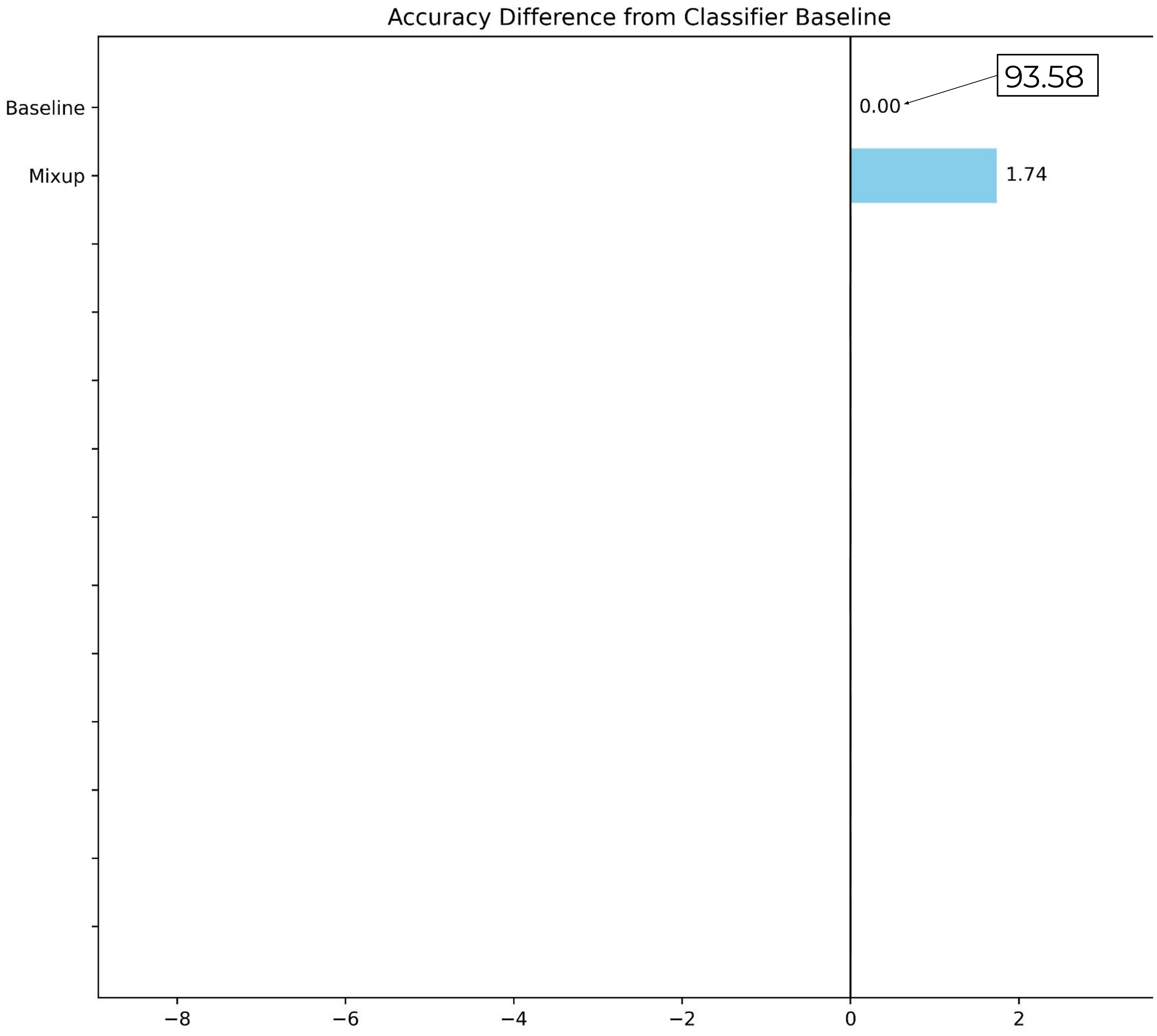
- Imbalance at class level
- Imbalance at domain level
- Disjoint domain
- Out-Of-Distribution classification

Iterations



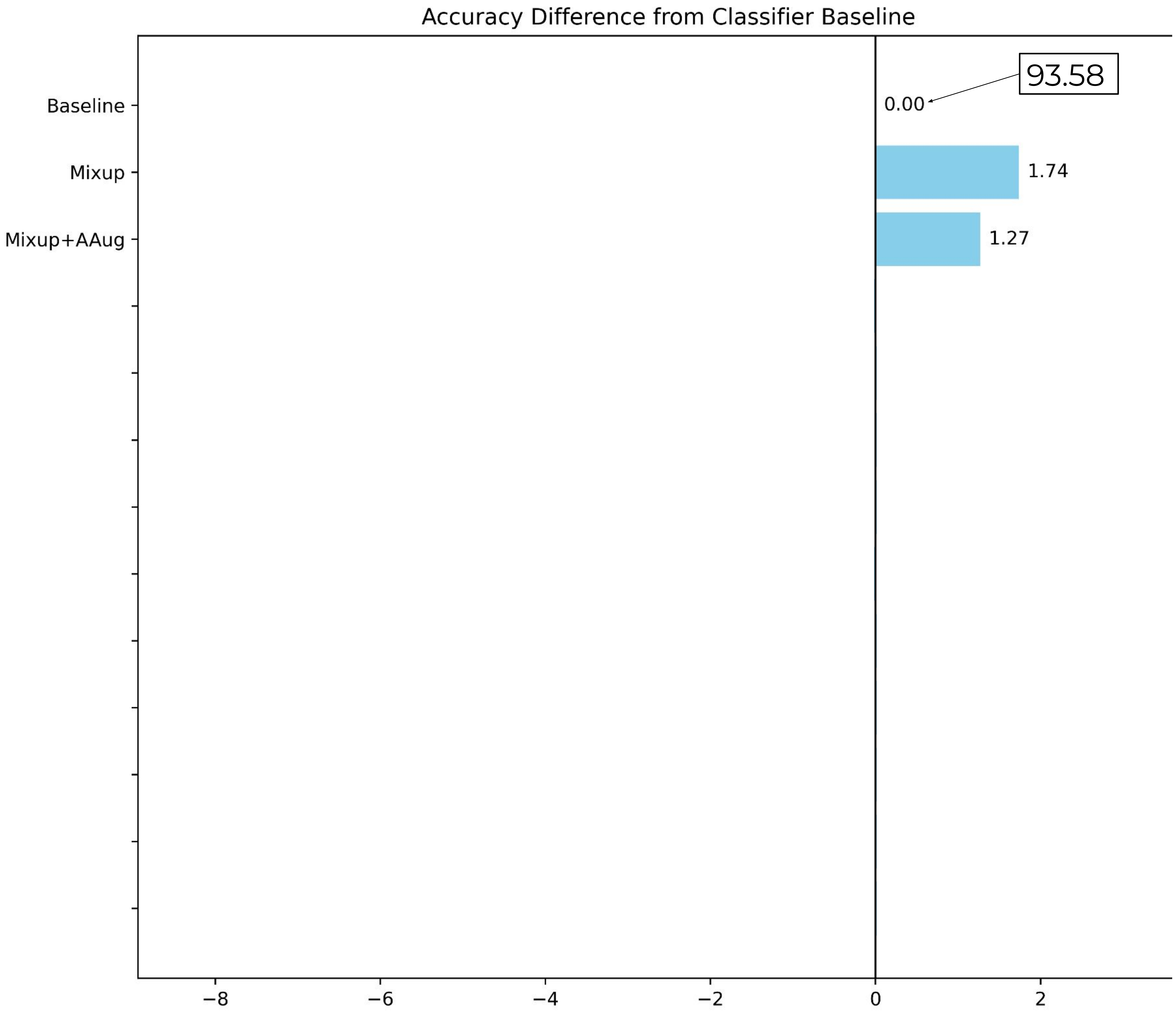


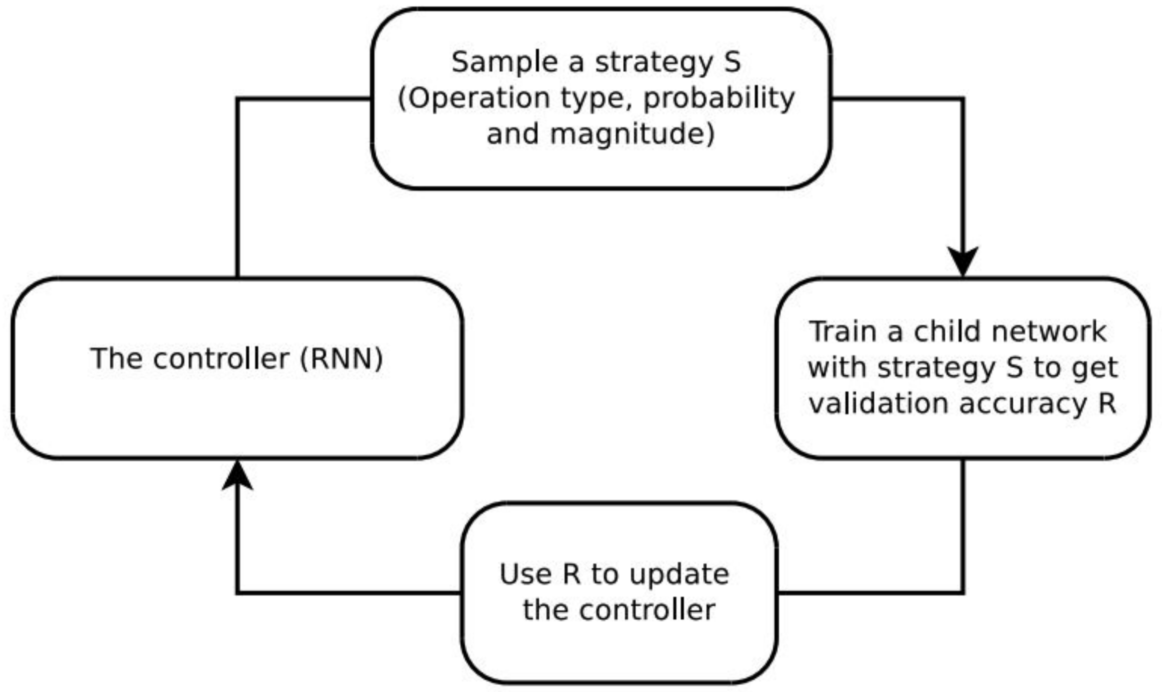
Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. Learning transferable visual models from natural language supervision. ICML 2021



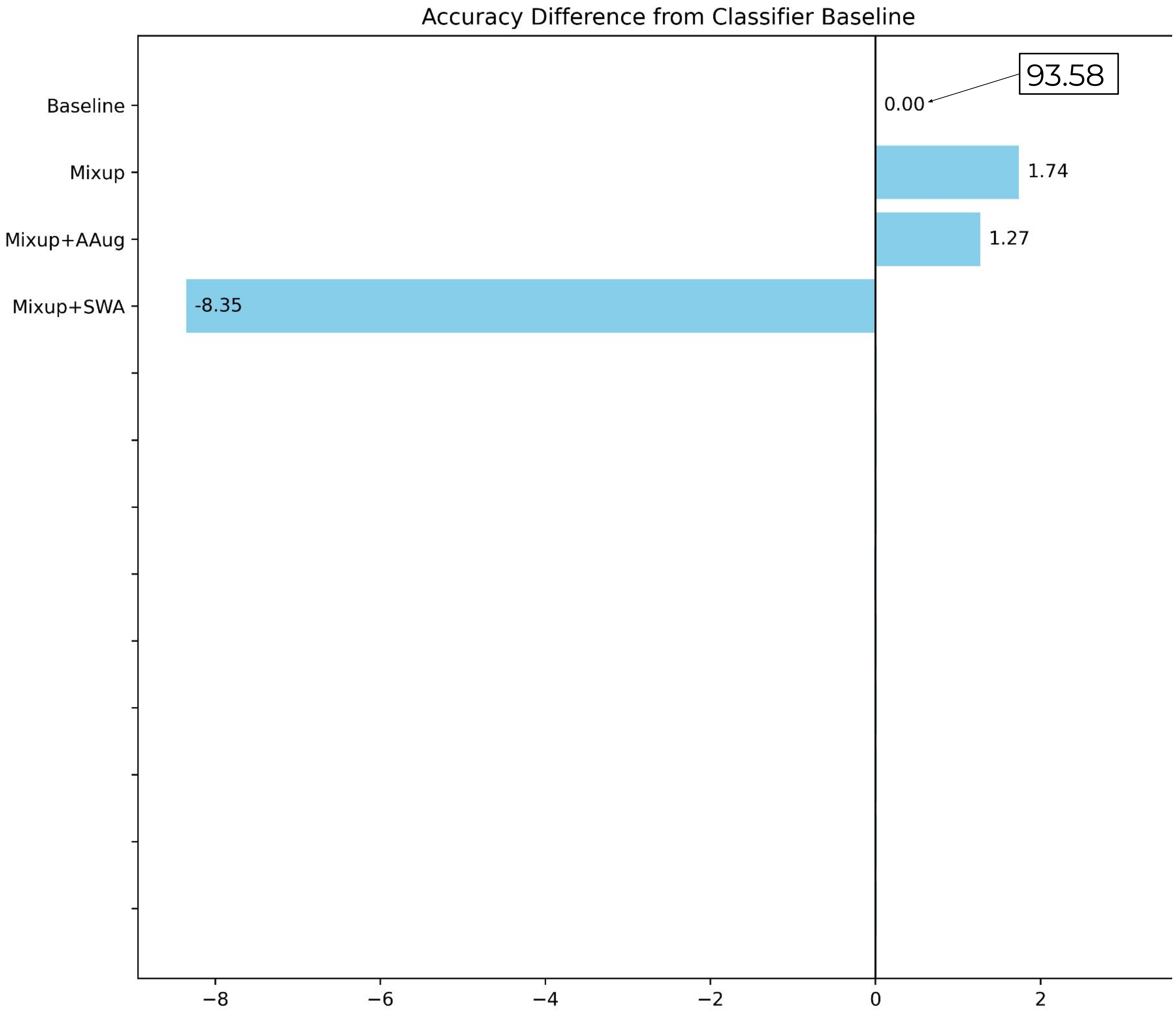

```python
# y1, y2 should be one-hot vectors
for (x1, y1), (x2, y2) in zip(loader1, loader2):
    lam = numpy.random.beta(alpha, alpha)
    x = Variable(lam * x1 + (1. - lam) * x2)
    y = Variable(lam * y1 + (1. - lam) * y2)
    optimizer.zero_grad()
    loss(net(x), y).backward()
    optimizer.step()
```

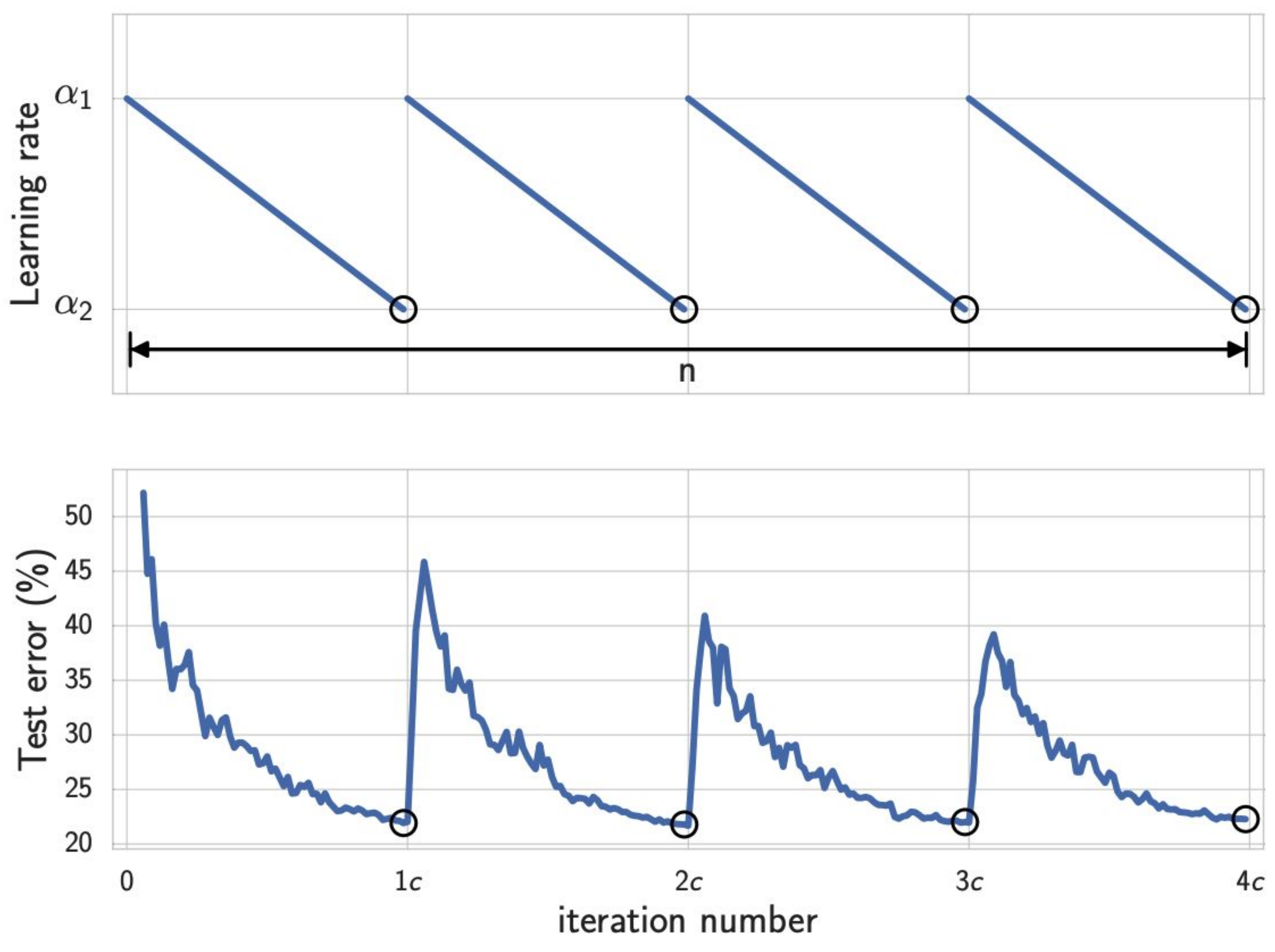
Zhang, Hongyi. mixup: Beyond empirical risk minimization. ICLR 2018

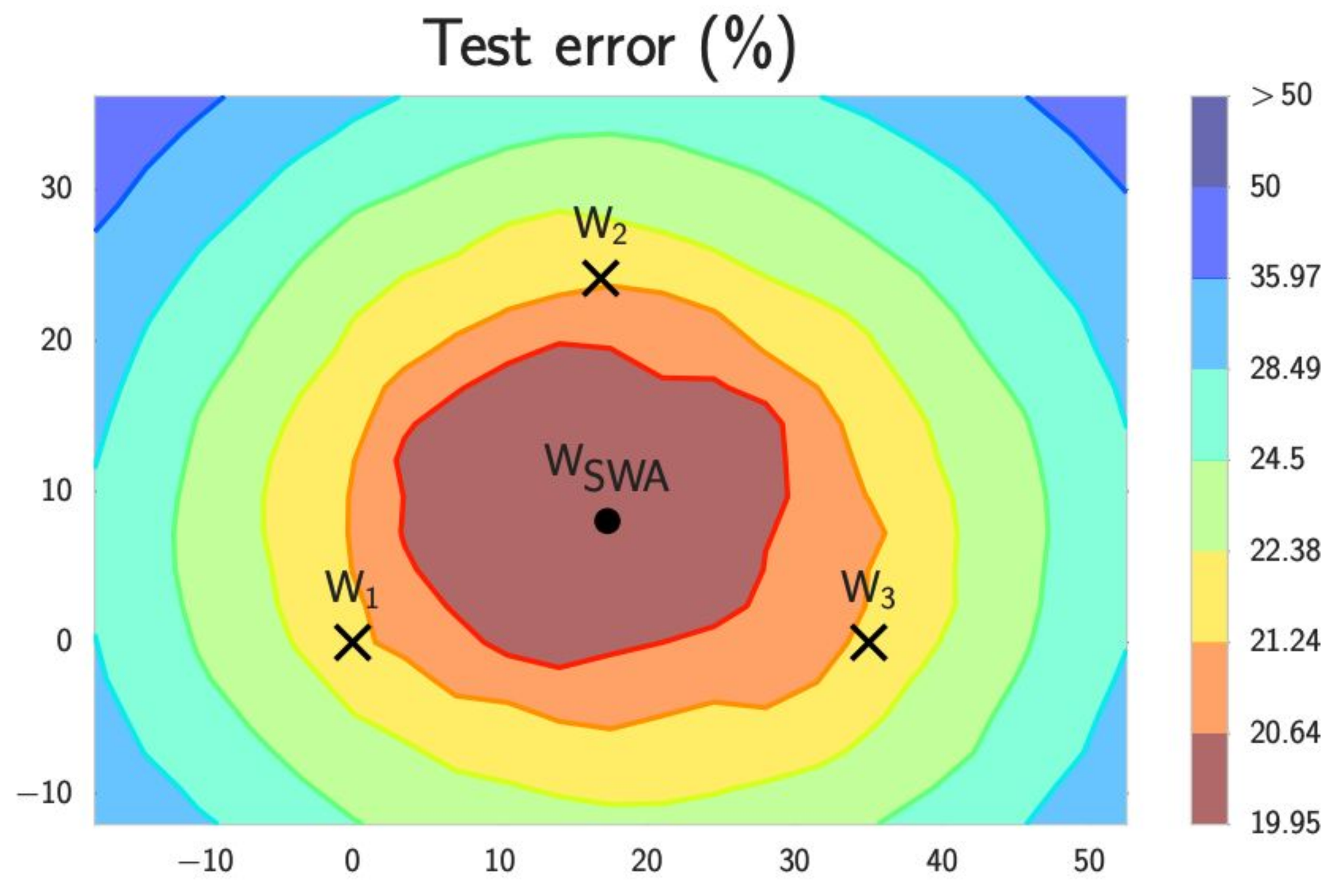




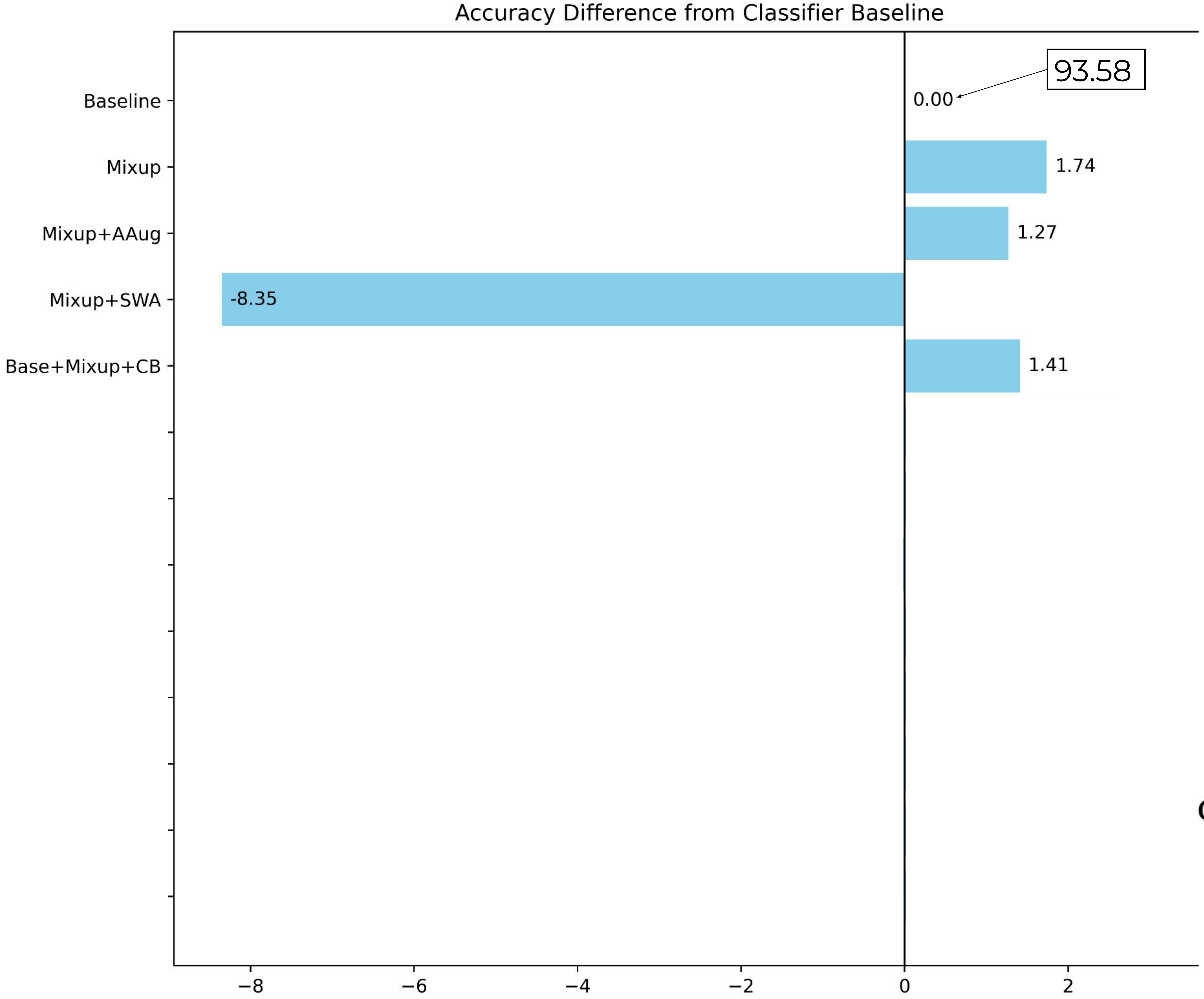
Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. Autoaugment: Learning augmentation policies from data. CVPR 2019.

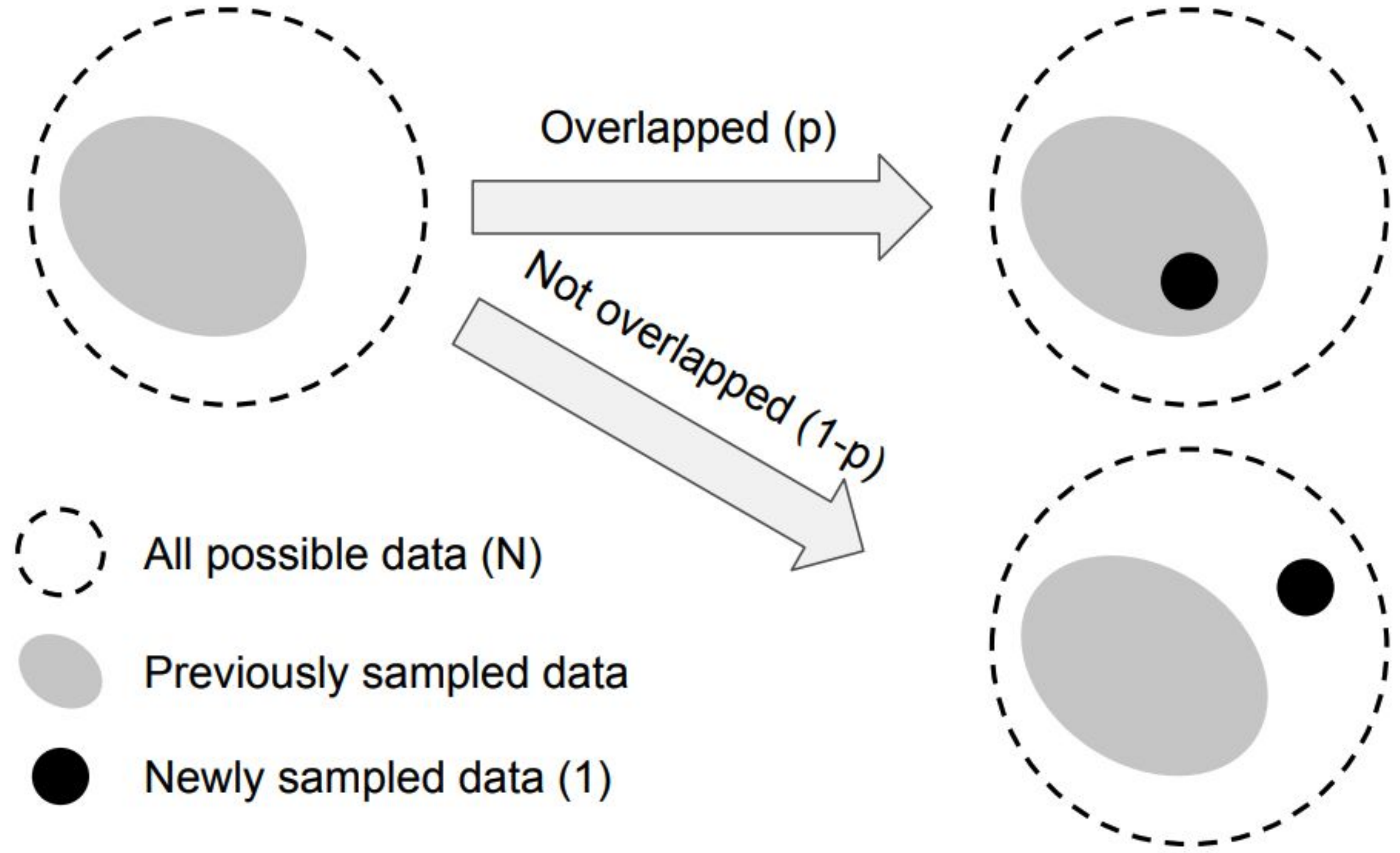


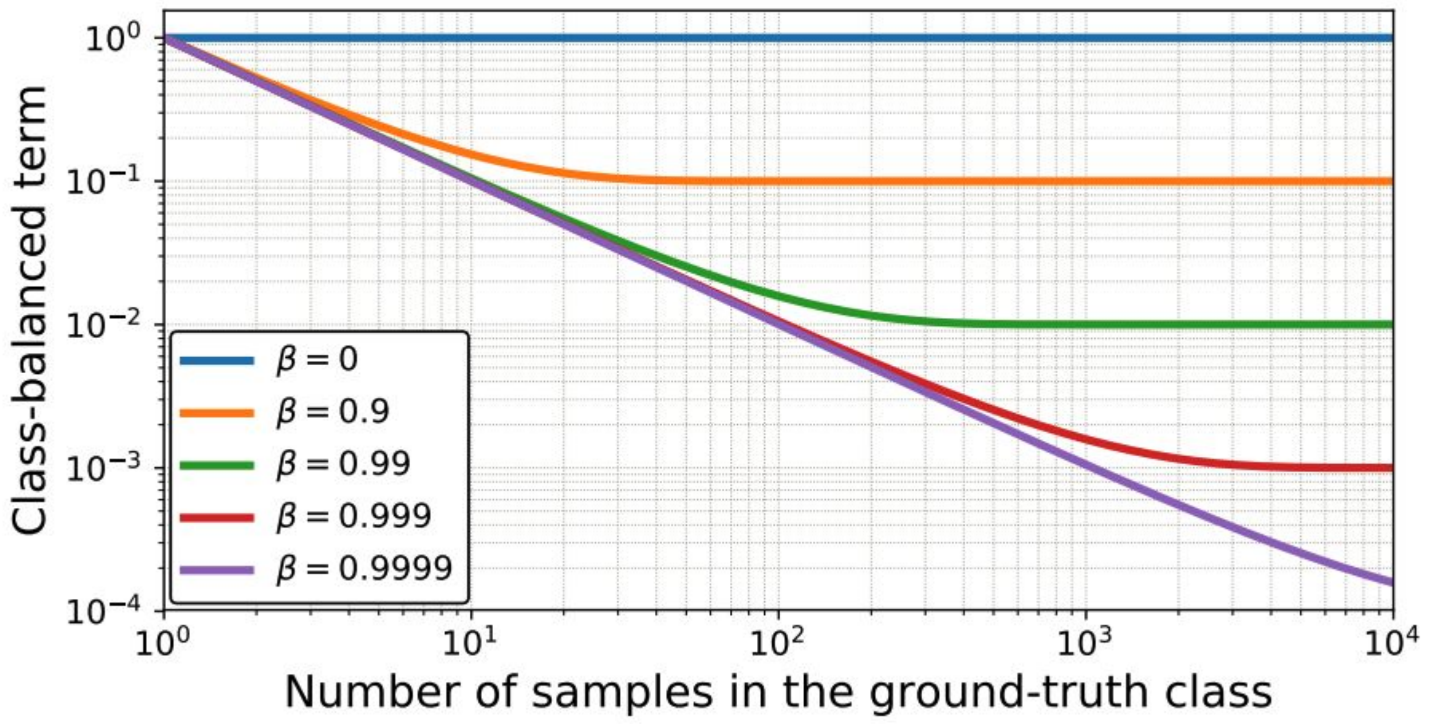




Izmailov, P., Podoprikhin, D., Garipov, T., Vetrov, D., & Wilson, A. G. Averaging weights leads to wider optima and better generalization. UAI 2018

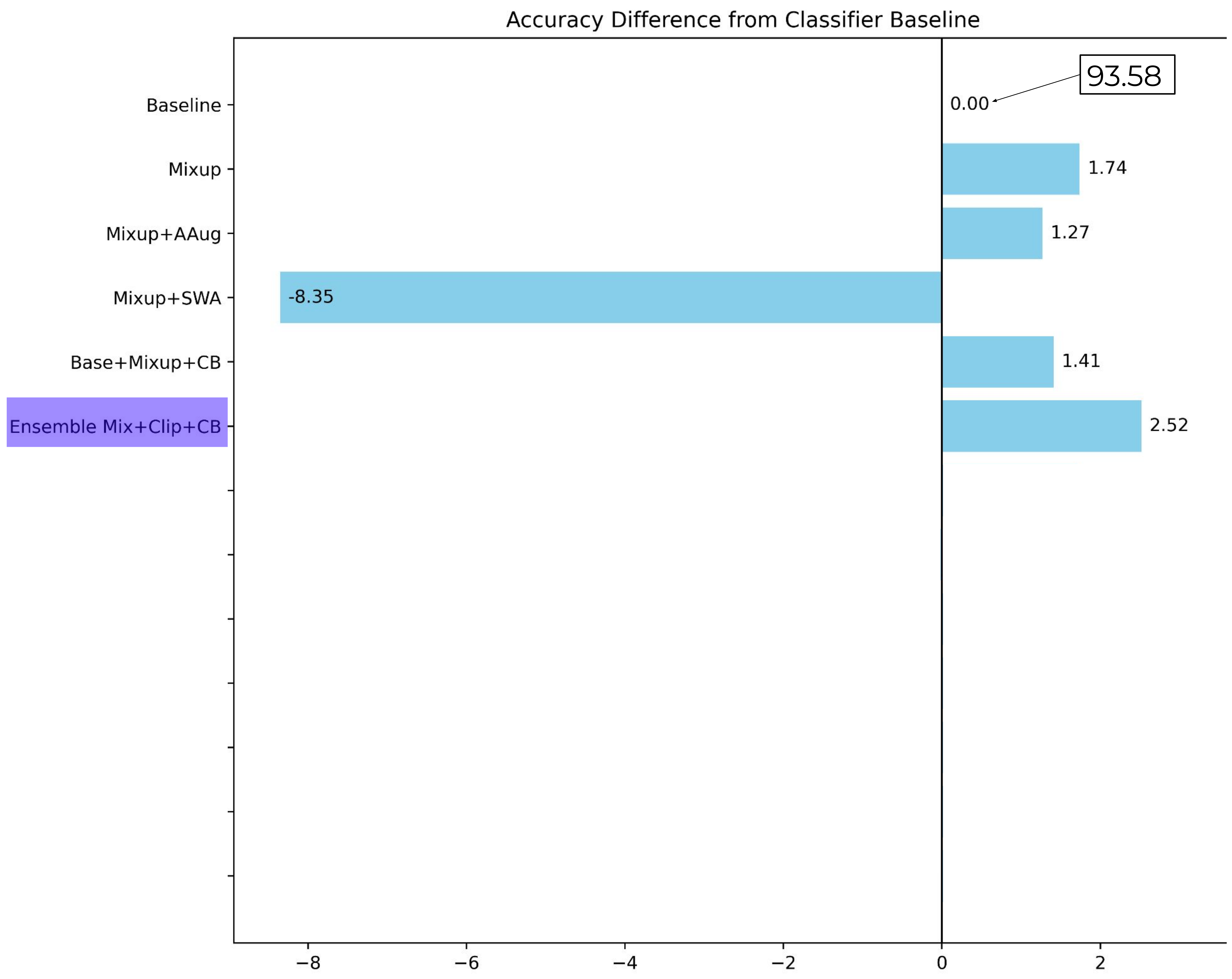


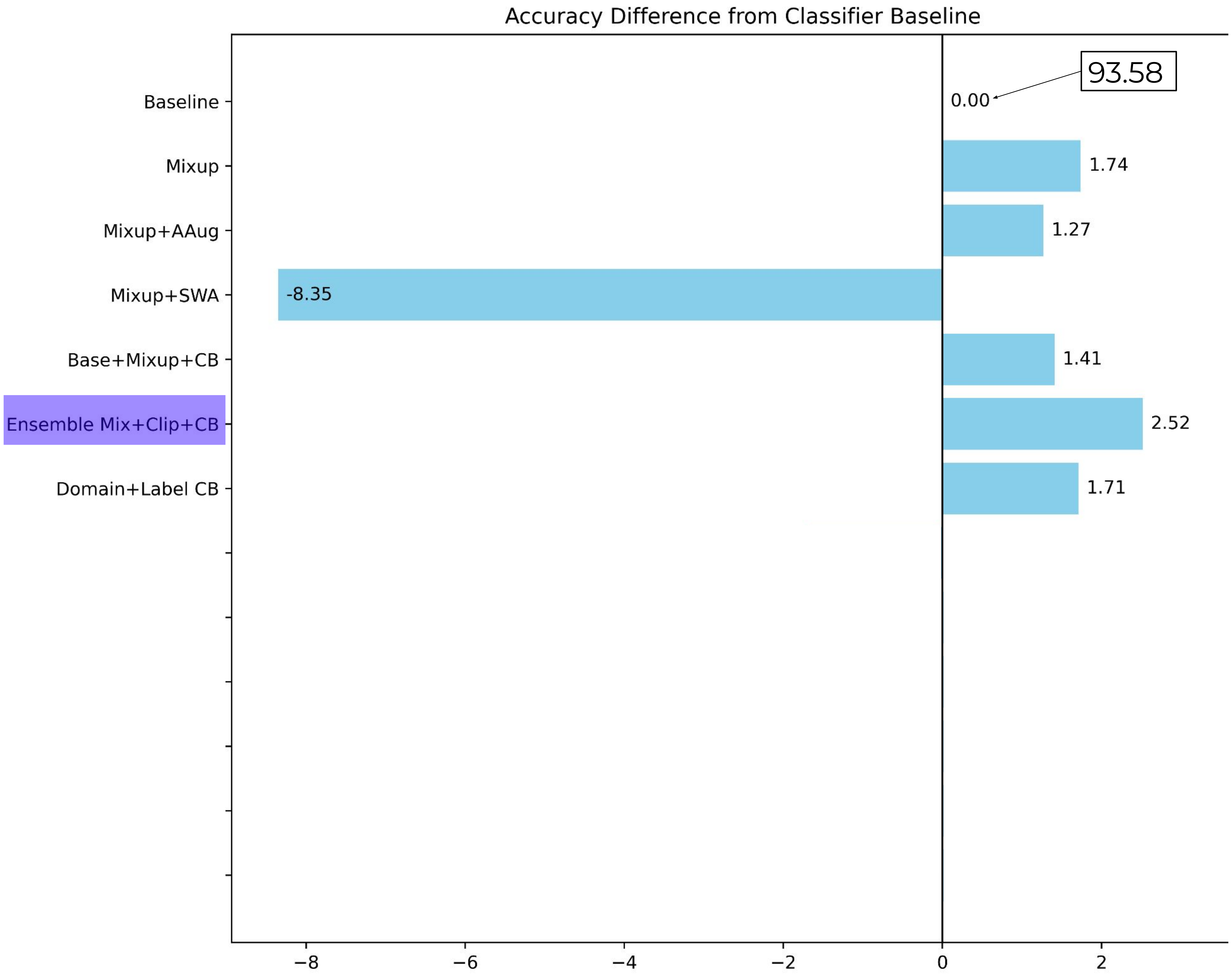


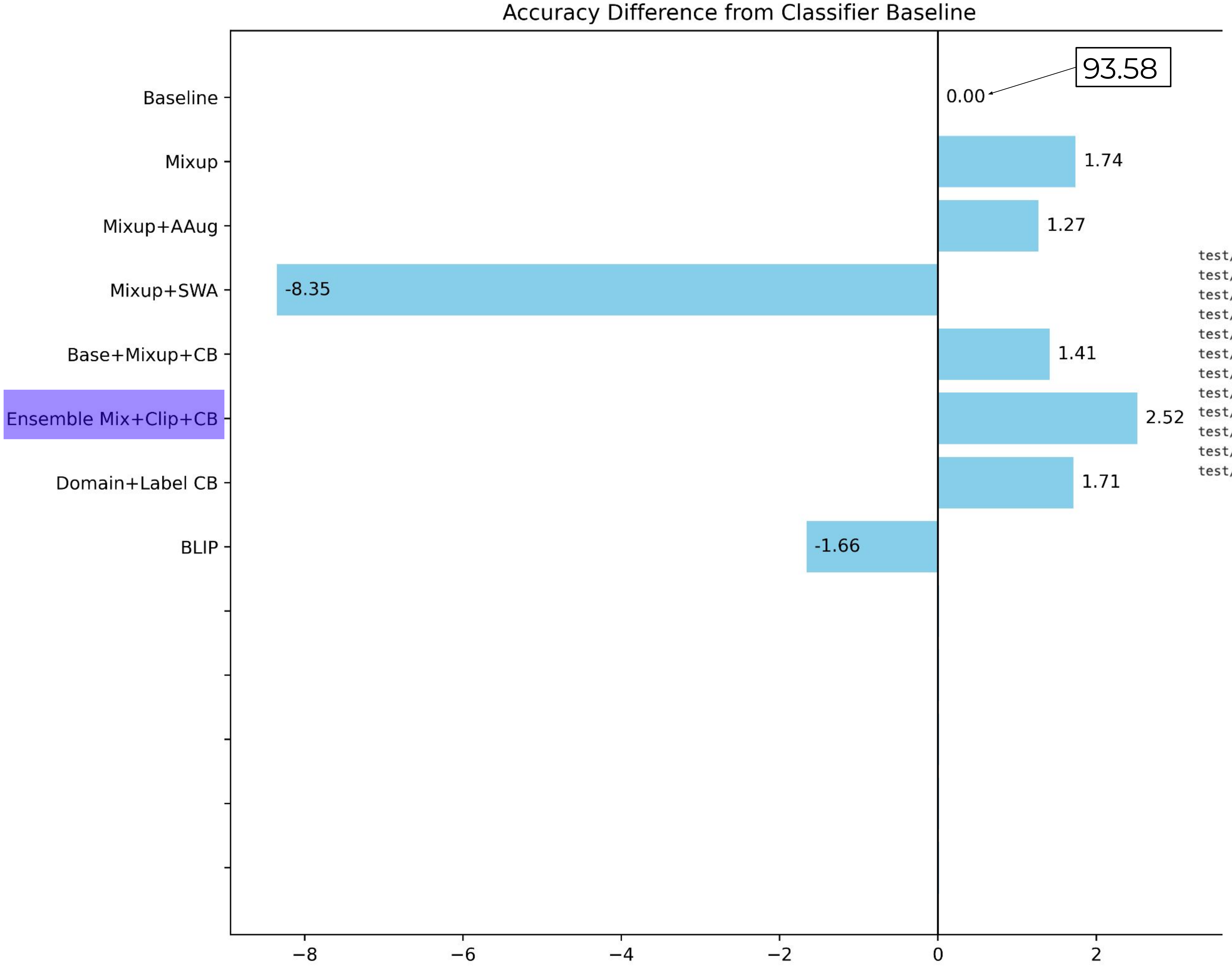


$$\mathbf{CB}_{\mathrm{softmax}}(\mathbf{z}, y) = -\frac{1-\beta}{1-\beta^{n_y}} \log \left(\frac{\exp(z_y)}{\sum_{j=1}^{C} \exp(z_j)} \right)$$

Cui, Y., Jia, M., Lin, T. Y., Song, Y., & Belongie, S. Class-balanced loss based on effective number of samples. CVPR 2019.

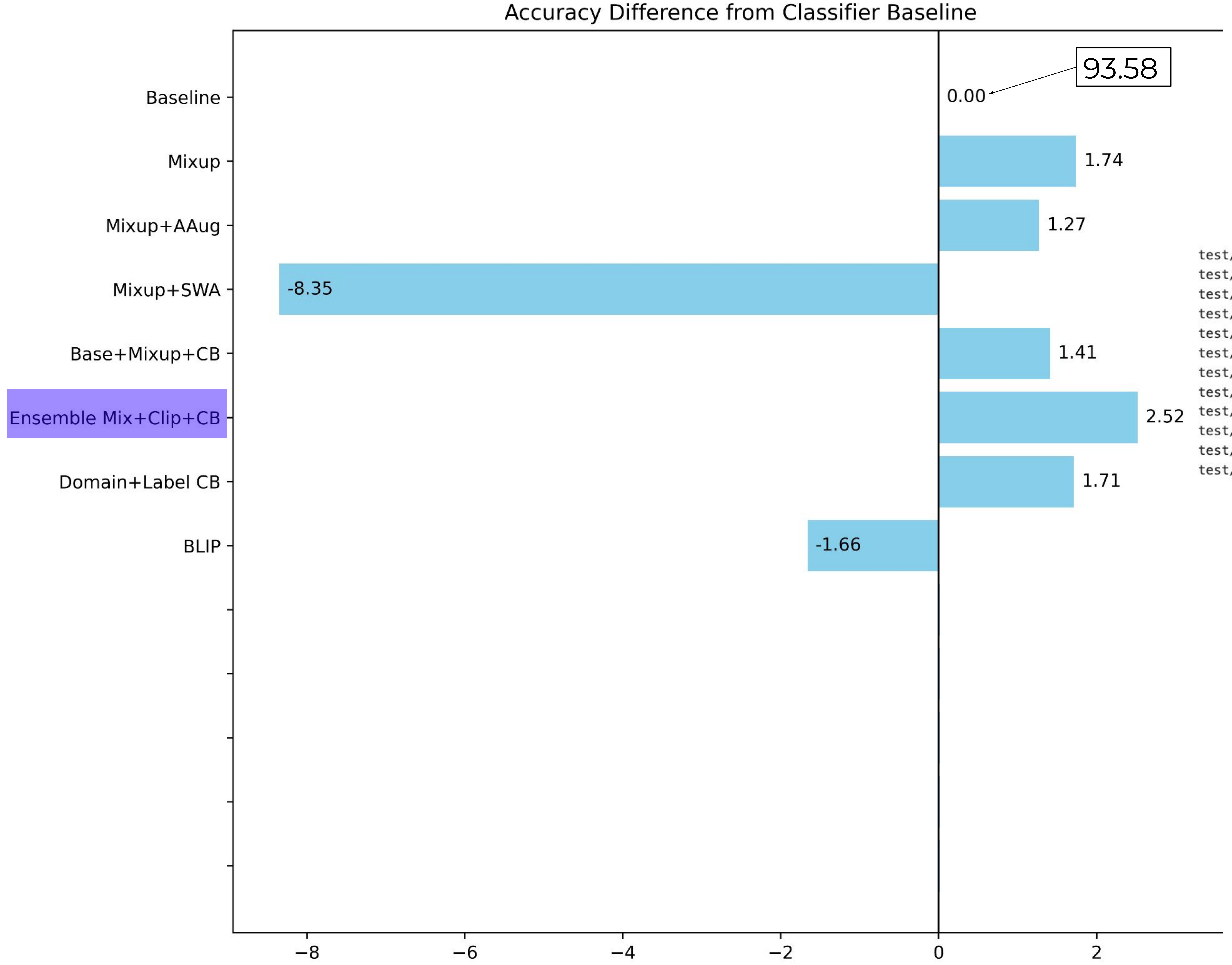

Accuracy Difference from Classifier Baseline
93.58
Baseline
0.00
Mixup
1.74
Mixup+AAug
1.27
Mixup+SWA
-8.35
Base+Mixup+CB
1.41
Ensemble Mix+Clip+CB
2.52
−8
−6
−4
−2
0
2


Accuracy Difference from Classifier Baseline
93.58
Baseline
0.00
Mixup
1.74
Mixup+AAug
1.27
Mixup+SWA
-8.35
Base+Mixup+CB
1.41
Ensemble Mix+Clip+CB
2.52
Domain+Label CB
1.71
−8
−6
−4
−2
0
2



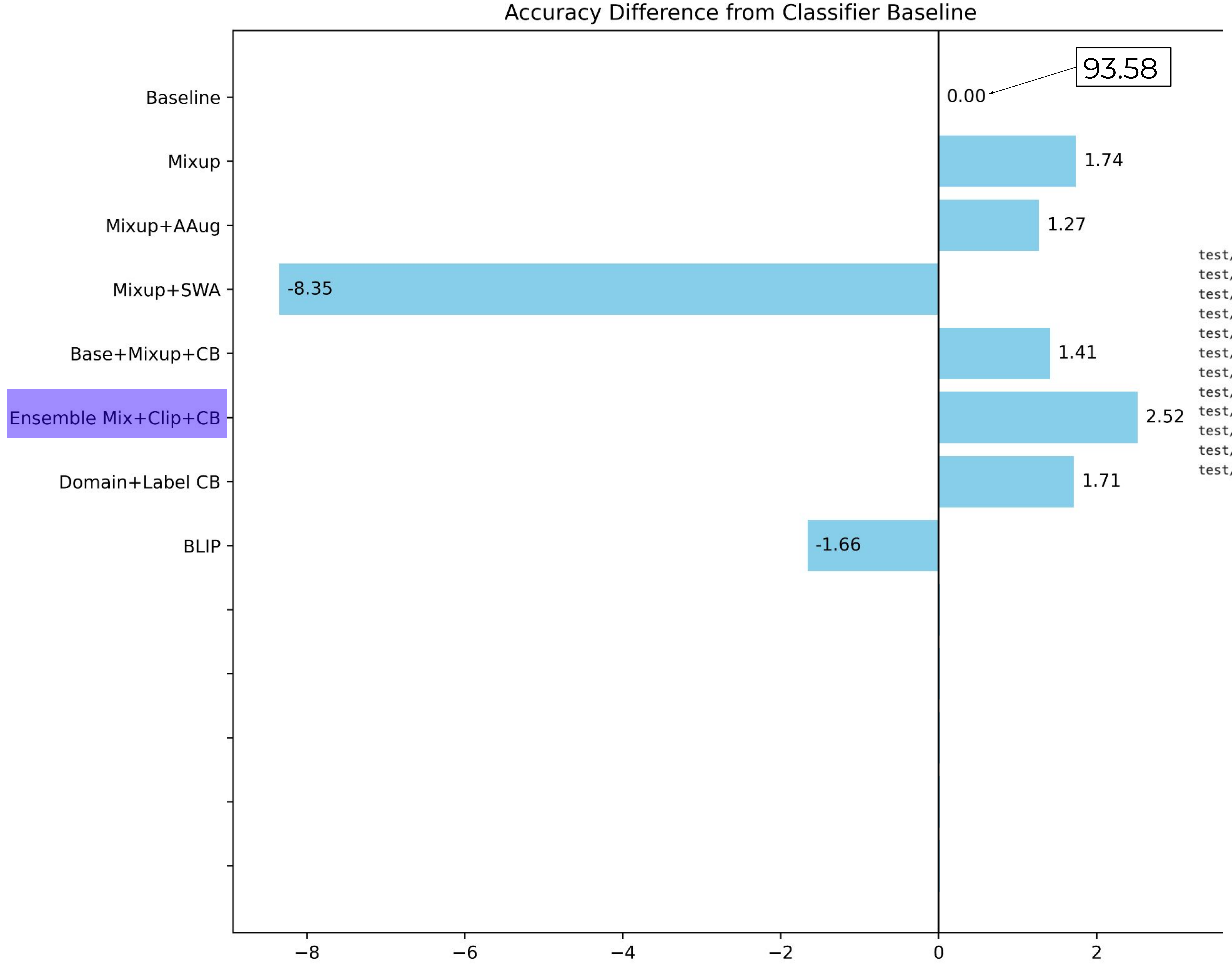

```
test/3GKVmKvzd3.png, a black and white drawing of a person on a surfboard
test/3HOKIG6B3x.png, a black and white drawing of a guitar
test/3Hiv44imeq.png, a black and white drawing of a fish
test/3HmGWsvmaa.png, a black and white drawing of a bee
test/3HohOAqJgS.png, a black and white drawing of a pen
test/3ICFwzNC4G.png, a black and white drawing of a table and chair
test/3InuzeSgX0.png, a black and white drawing of a leaf
test/3JMcjFWwur.png, a black and white drawing of a deer
test/3KkjsBGt0I.png, a black and white drawing of a train
test/3Klc0XABkX.png, a black and white outline of a thumb up
test/3LSIe659UL.png, a black and white drawing of a bag
test/3LfIcoZdPI.png, a black and white drawing of a fish
```

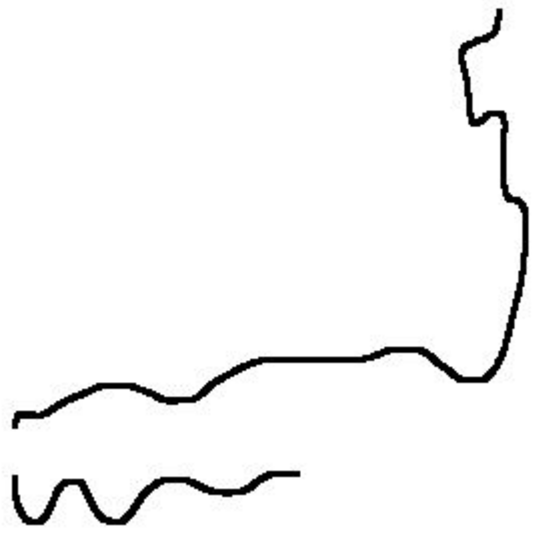
Li, J., Li, D., Xiong, C., & Hoi, S. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. ICML 2022



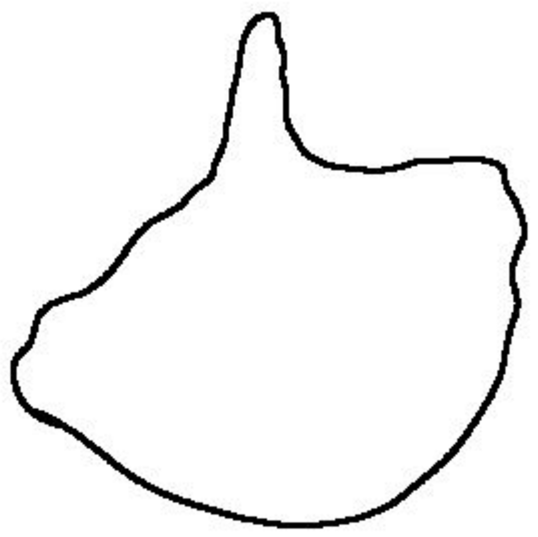

```
test/3GKVmKvzd3.png, a black and white drawing of a person on a surfboard
test/3HOKIG6B3x.png, a black and white drawing of a guitar
test/3Hiv44imeq.png, a black and white drawing of a fish
test/3HmGWsvmaa.png, a black and white drawing of a bee
test/3HohOAqJgS.png, a black and white drawing of a pen
test/3ICFwzNC4G.png, a black and white drawing of a table and chair
test/3InuzeSgX0.png, a black and white drawing of a leaf
test/3JMcjFWwur.png, a black and white drawing of a deer
test/3KkjsBGt0I.png, a black and white drawing of a train
test/3Klc0XABkX.png, a black and white outline of a thumb up
test/3LSIe659UL.png, a black and white drawing of a bag
test/3LfIcoZdPI.png, a black and white drawing of a fish
```

Li, J., Li, D., Xiong, C., & Hoi, S. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. ICML 2022

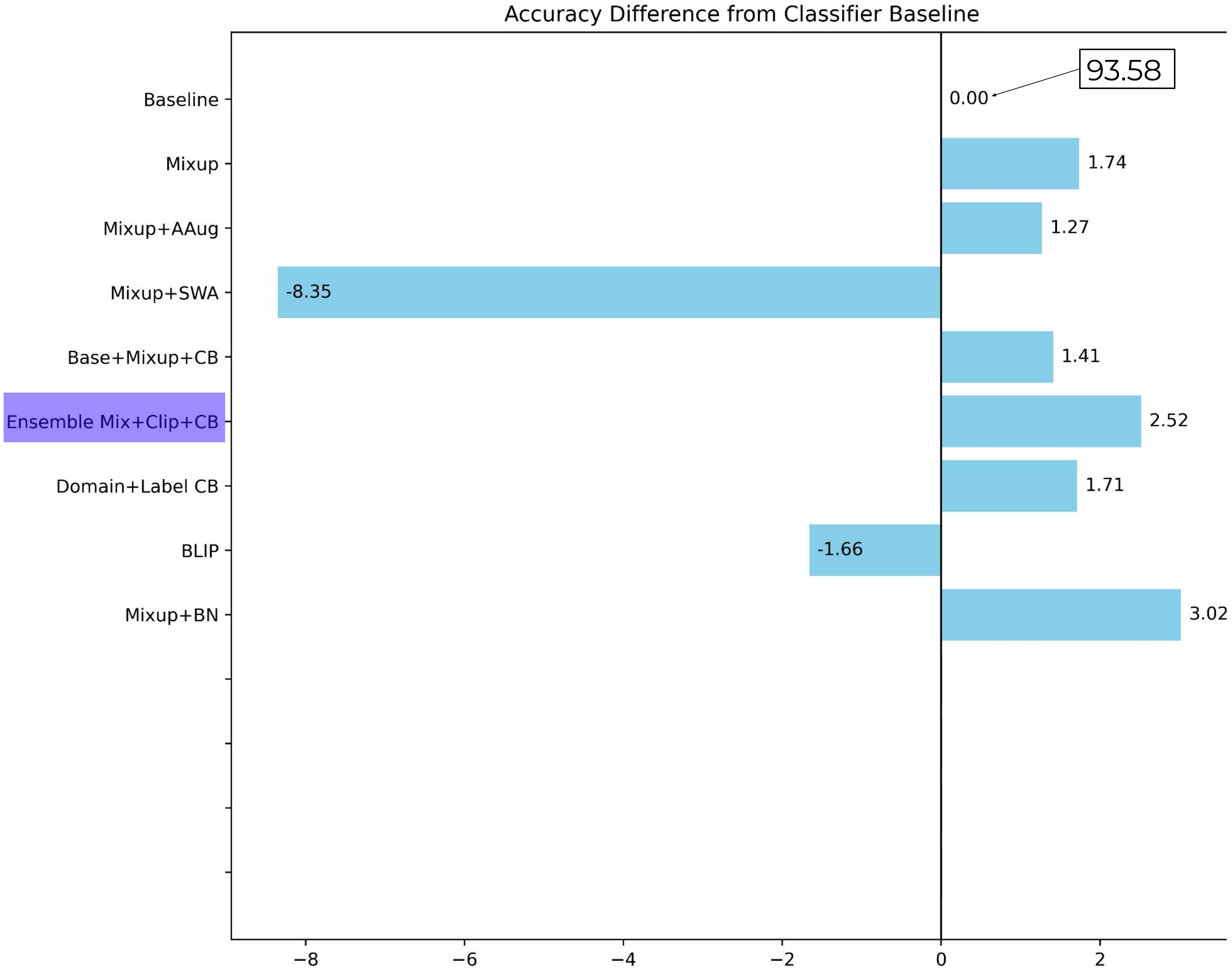




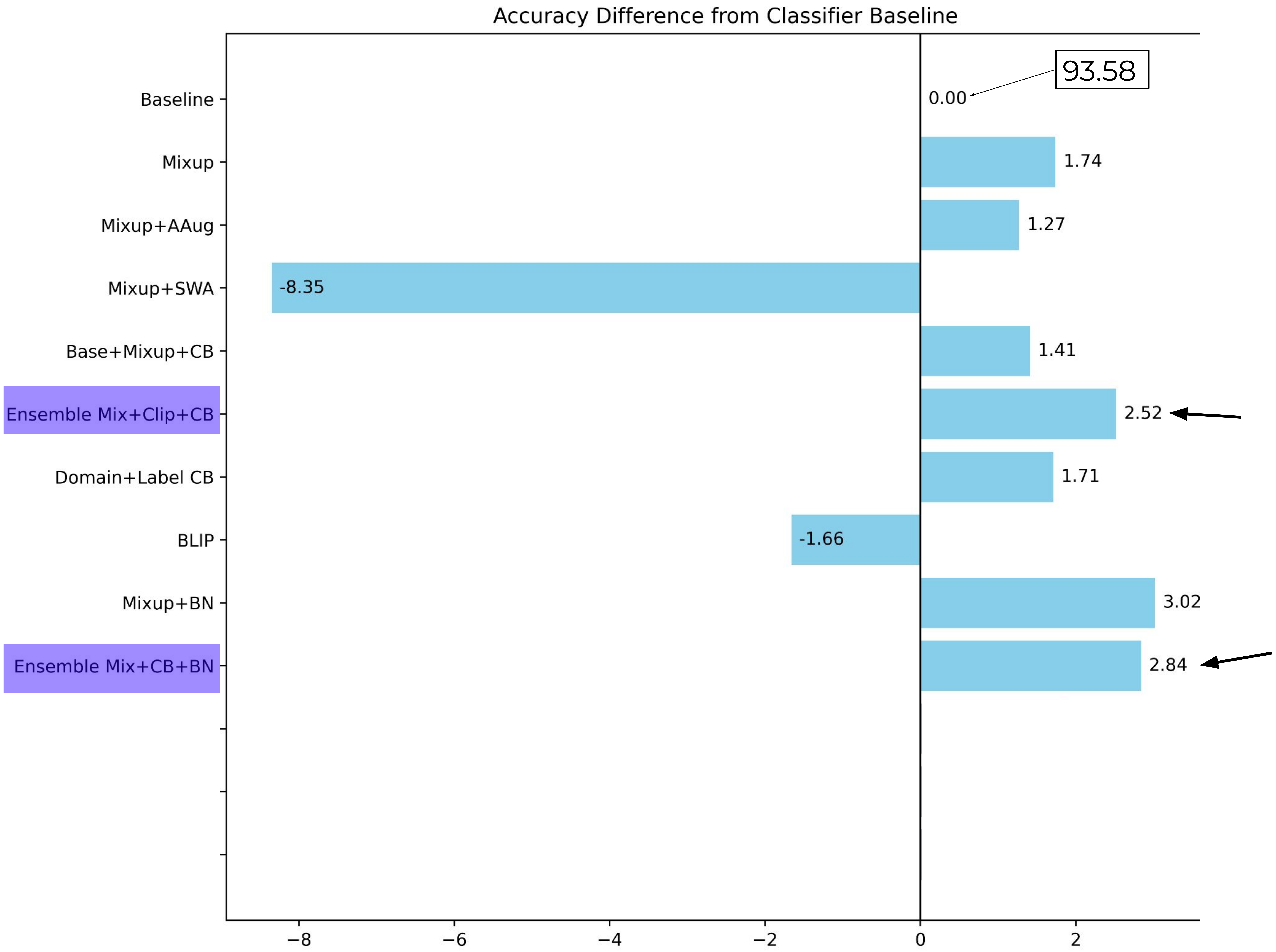
```
test/3GKVmKvzd3.png, a black and white drawing of a person on a surfboard
test/3HOKIG6B3x.png, a black and white drawing of a guitar
test/3Hiv44imeq.png, a black and white drawing of a fish
test/3HmGWsvmaa.png, a black and white drawing of a bee
test/3HohOAqJgS.png, a black and white drawing of a pen
test/3ICFwzNC4G.png, a black and white drawing of a table and chair
test/3InuzeSgX0.png, a black and white drawing of a leaf
test/3JMcjFWwur.png, a black and white drawing of a deer
test/3KkjsBGt0I.png, a black and white drawing of a train
test/3Klc0XABkX.png, a black and white outline of a thumb up
test/3LSIe659UL.png, a black and white drawing of a bag
test/3LfIcoZdPI.png, a black and white drawing of a fish
```



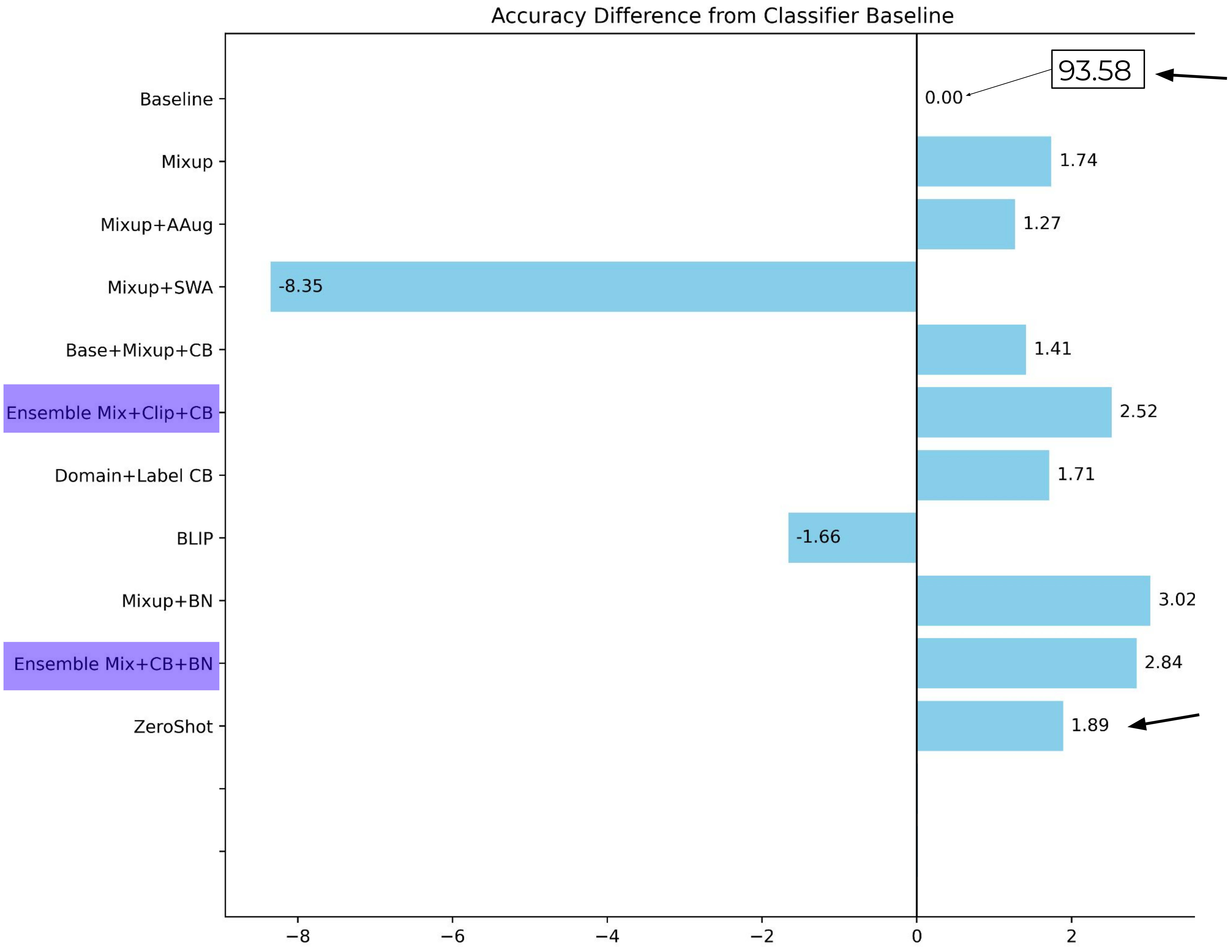
Li, J., Li, D., Xiong, C., & Hoi, S. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. ICML 2022



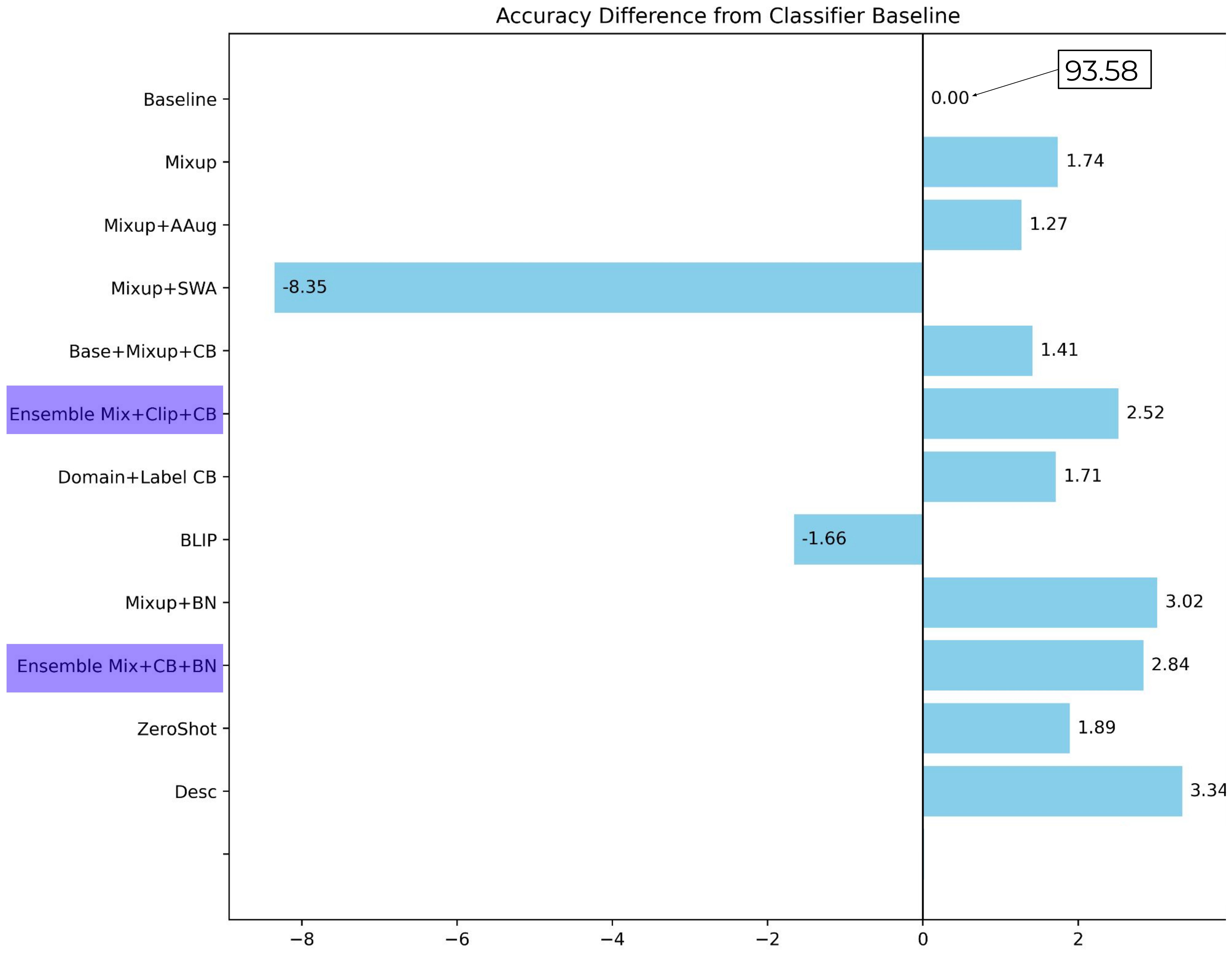
Ioffe, S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. ICML 2015

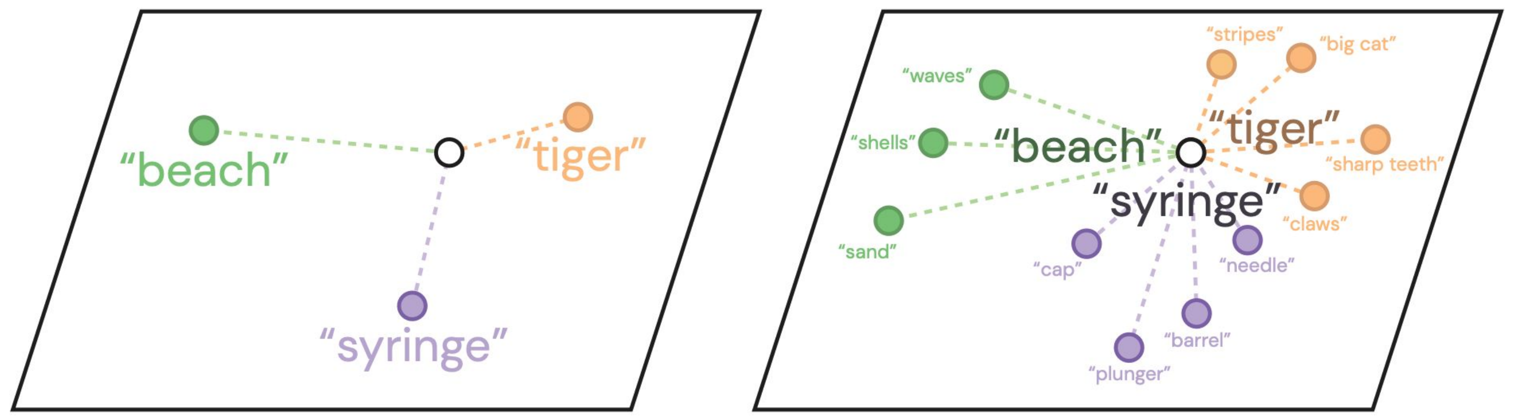


Accuracy Difference from Classifier Baseline
93.58
Baseline
0.00
Mixup
1.74
Mixup+AAug
1.27
Mixup+SWA
-8.35
Base+Mixup+CB
1.41
Ensemble Mix+Clip+CB
2.52
Domain+Label CB
1.71
BLIP
-1.66
Mixup+BN
3.02
Ensemble Mix+CB+BN
2.84
−8
−6
−4
−2
0
2

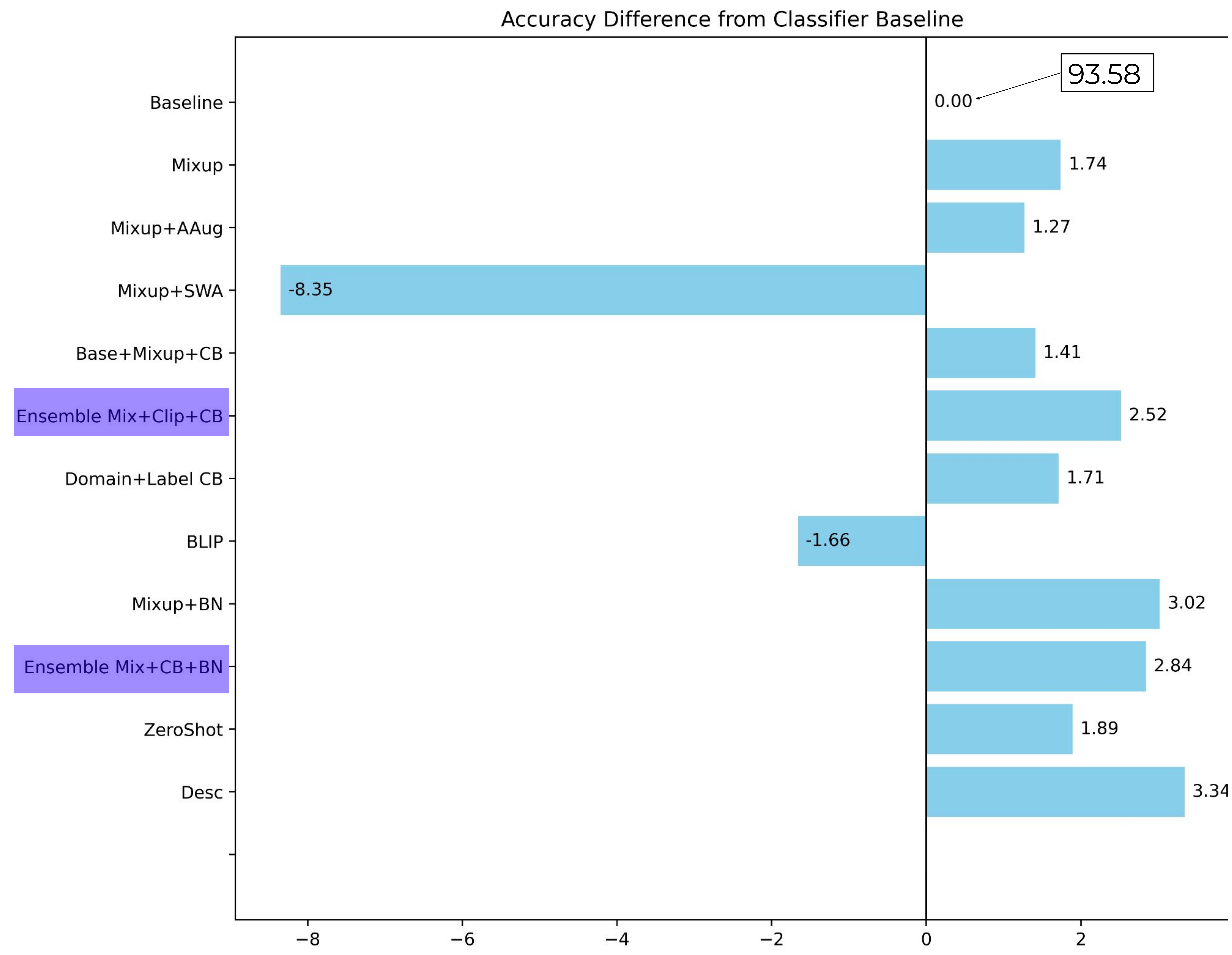


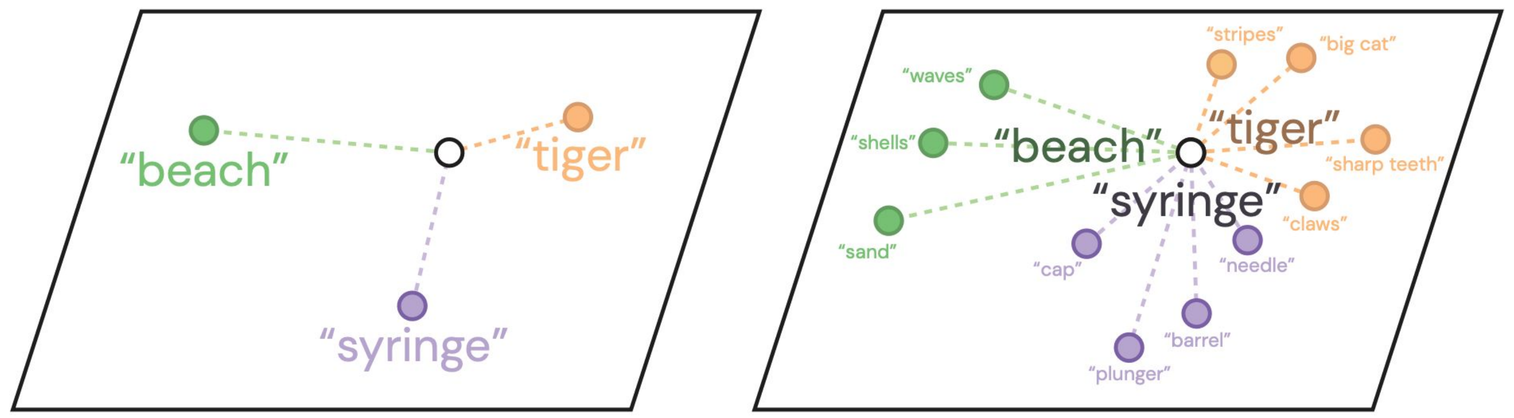
Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. Learning transferable visual models from natural language supervision. ICML 2021





Menon, S., & Vondrick, C. Visual classification via description from large language models. ICLR 2023. Notable Top 5% (Oral)

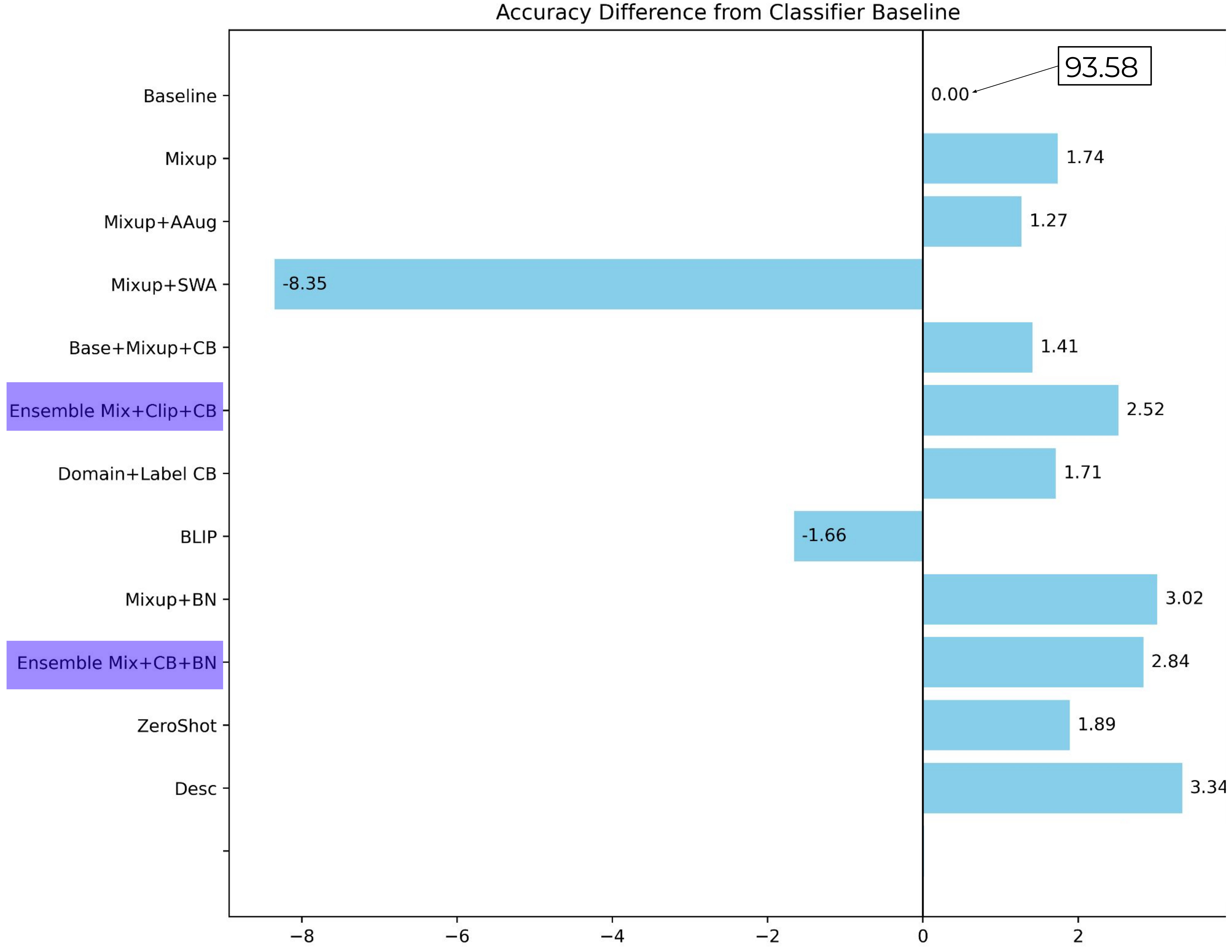


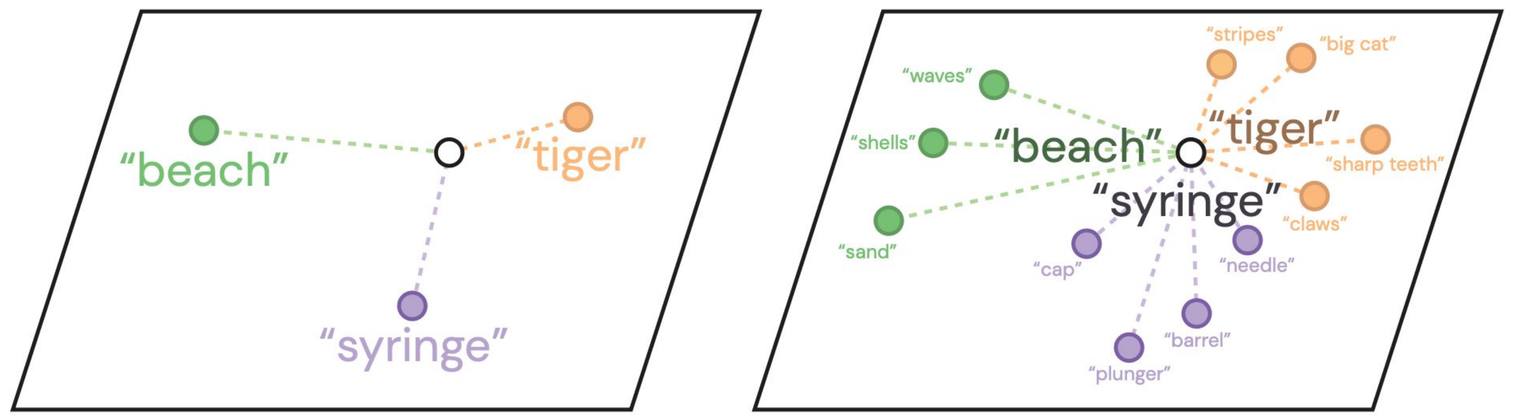


Violin
- a stringed instrument
- typically has four strings
- a wooden body
- a neck and fingerboard
- tuning pegs
- a bridge
- a soundpost
- f-holes
- a bow

Pirate ship
- a large, sailing vessel
- a flag with a skull and crossbones
- cannons on the deck
- a wooden hull
- portholes
- rigging
- a crow's nest

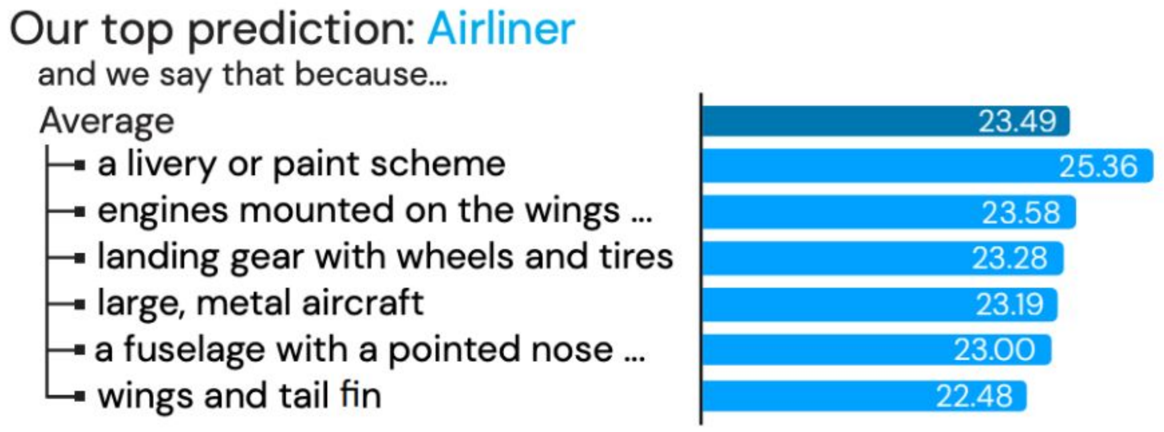
Menon, S., & Vondrick, C. Visual classification via description from large language models. ICLR 2023. Notable Top 5% (Oral)



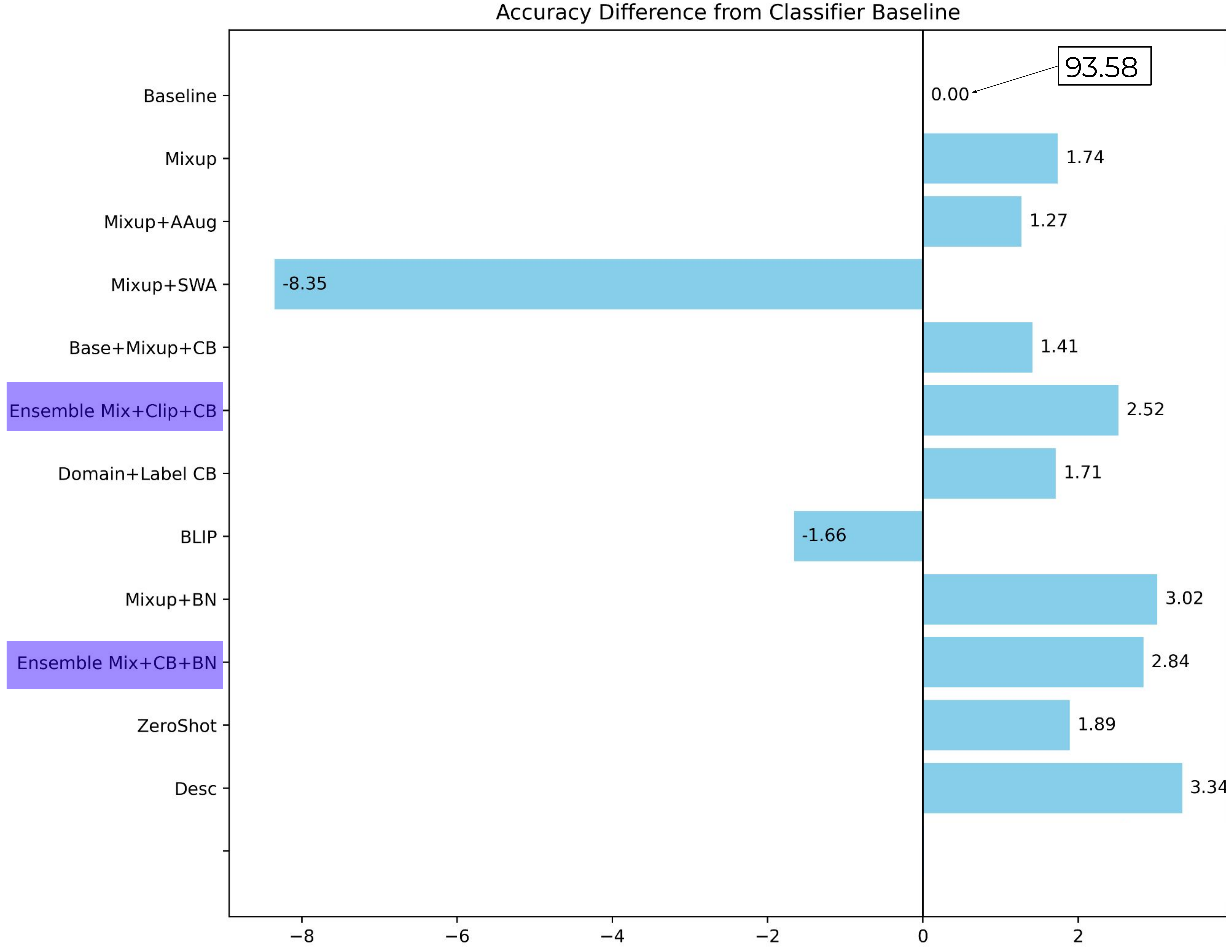


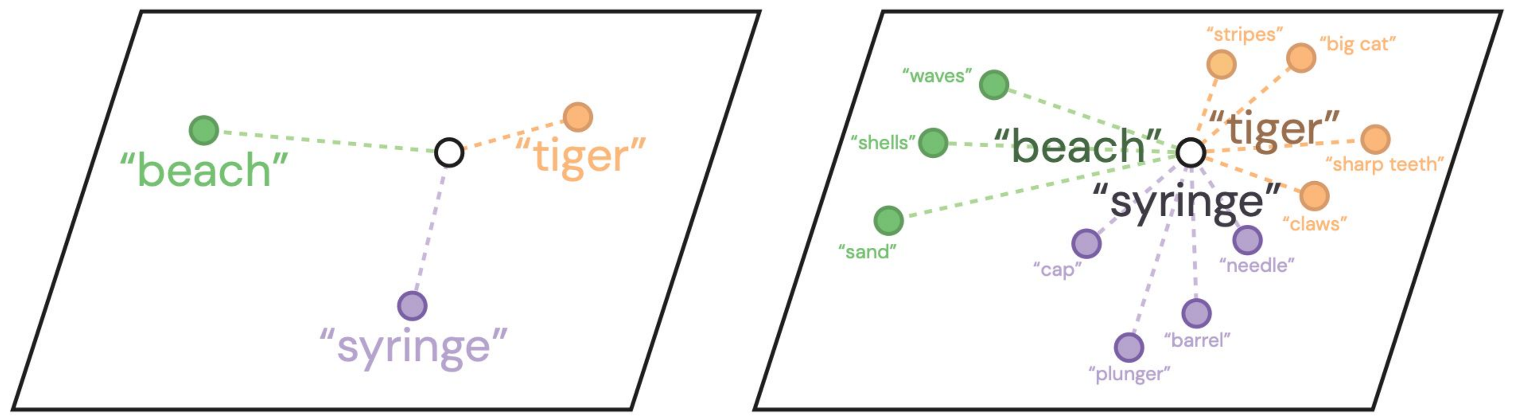
Violin
- a stringed instrument
- typically has four strings
- a wooden body
- a neck and fingerboard
- tuning pegs
- a bridge
- a soundpost
- f-holes
- a bow

Pirate ship
- a large, sailing vessel
- a flag with a skull and crossbones
- cannons on the deck
- a wooden hull
- portholes
- rigging
- a crow's nest



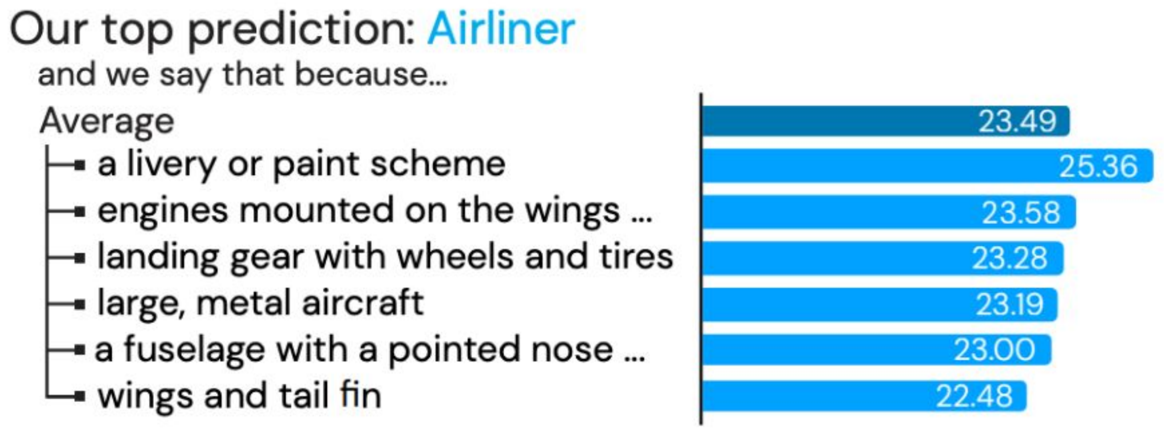
Menon, S., & Vondrick, C. Visual classification via description from large language models. ICLR 2023. Notable Top 5% (Oral)



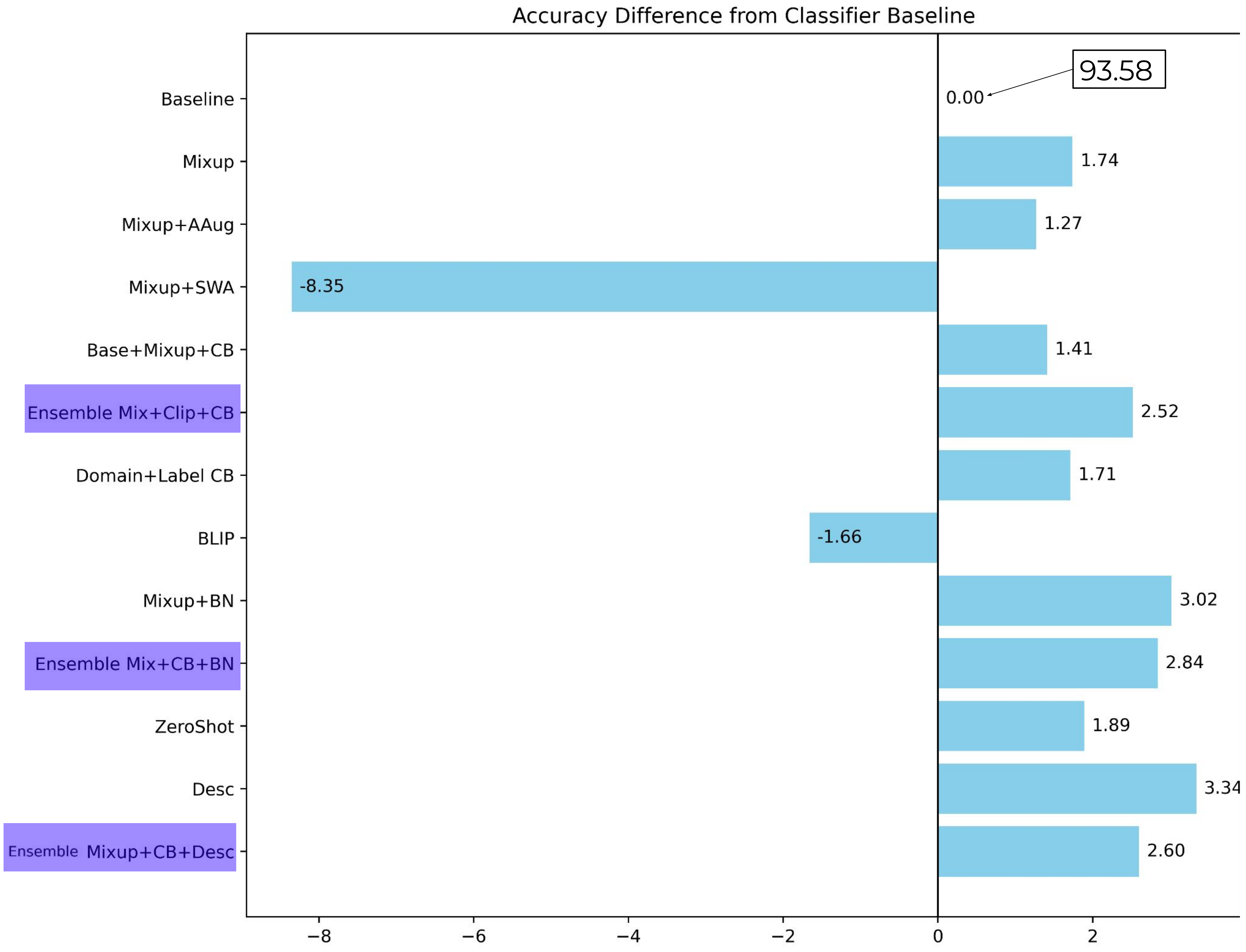


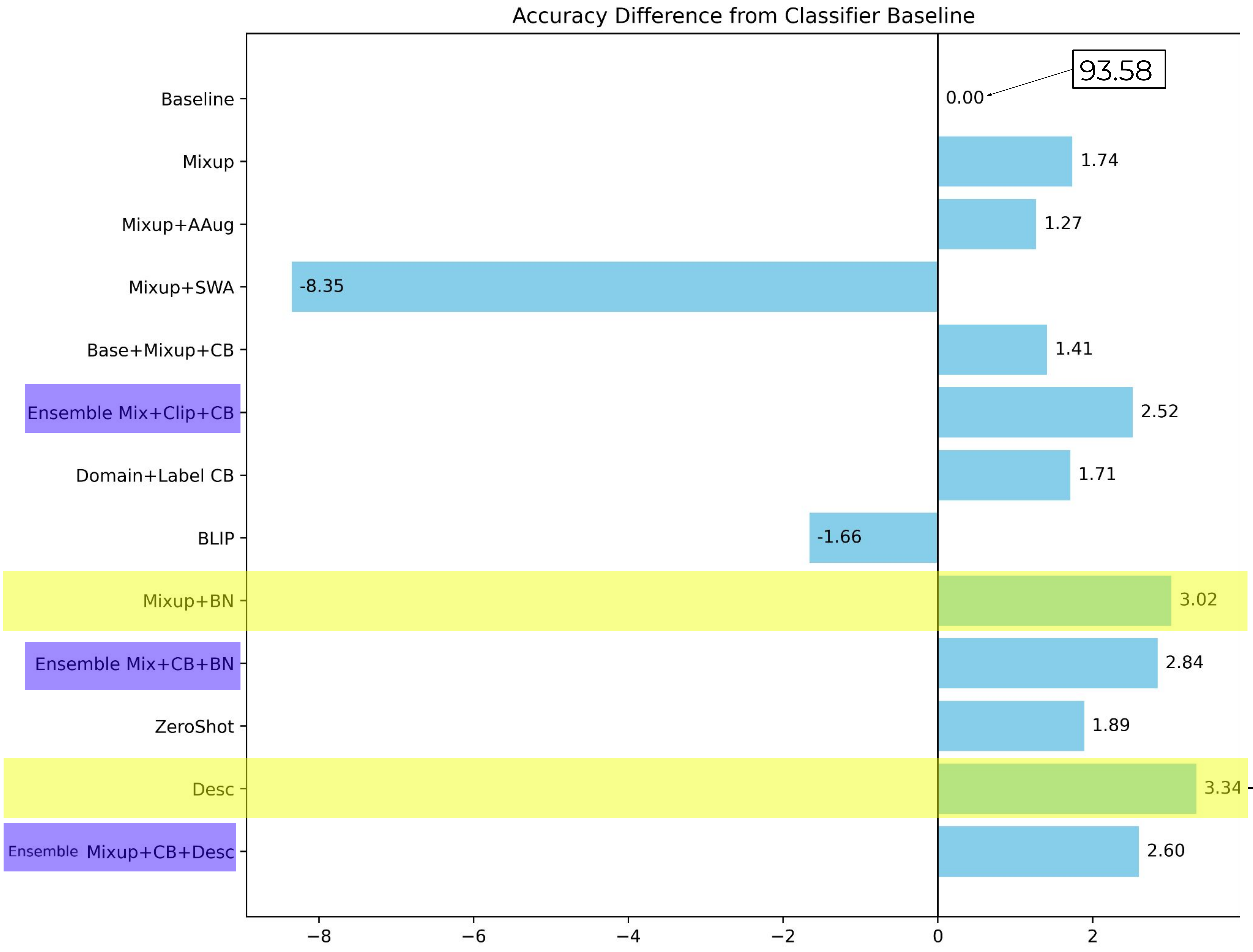
Violin
- a stringed instrument
- typically has four strings
- a wooden body
- a neck and fingerboard
- tuning pegs
- a bridge
- a soundpost
- f-holes
- a bow

Pirate ship
- a large, sailing vessel
- a flag with a skull and crossbones
- cannons on the deck
- a wooden hull
- portholes
- rigging
- a crow's nest



Menon, S., & Vondrick, C. Visual classification via description from large language models. ICLR 2023. Notable Top 5% (Oral)


Accuracy Difference from Classifier Baseline
93.58
Baseline
0.00
Mixup
1.74
Mixup+AAug
1.27
Mixup+SWA
-8.35
Base+Mixup+CB
1.41
Ensemble Mix+Clip+CB
2.52
Domain+Label CB
1.71
BLIP
-1.66
Mixup+BN
3.02
Ensemble Mix+CB+BN
2.84
ZeroShot
1.89
Desc
3.34
Ensemble Mixup+CB+Desc
2.60
−8
−6
−4
−2
0
2



#	△	Team	Members	Score
1	—	ThisIsTrifficult		0.53700

Takeaways

- Embracing foundational models is essential.
- Foundational models behave differently compared to others.
- There is a gap between what works on paper and what actually works.
- Simple tricks go a long way in improving accuracy.
- If simply adding losses did not help, ensemble them.

Citations

- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. Learning transferable visual models from natural language supervision. ICML 2021
- Zhang, Hongyi. mixup: Beyond empirical risk minimization. ICLR 2018
- Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. Autoaugment: Learning augmentation policies from data. CVPR 2019.
- Izmailov, P., Podoprikhin, D., Garipov, T., Vetrov, D., & Wilson, A. G. Averaging weights leads to wider optima and better generalization. UAI 2018
- Cui, Y., Jia, M., Lin, T. Y., Song, Y., & Belongie, S. Class-balanced loss based on effective number of samples. CVPR 2019.
- Li, J., Li, D., Xiong, C., & Hoi, S. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. ICML 2022
- Ioffe, S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. ICML 2015

- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. Learning transferable visual models from natural language supervision. ICML 2021

- Menon, S., & Vondrick, C. Visual classification via description from large language models. ICLR 2023.

Rahul Vigneswaran K

Thank you so much to the TAs for tolerating me ❤️