BOSCH MODEL EXTRACTION

Team 6

KEY CHALLENGES

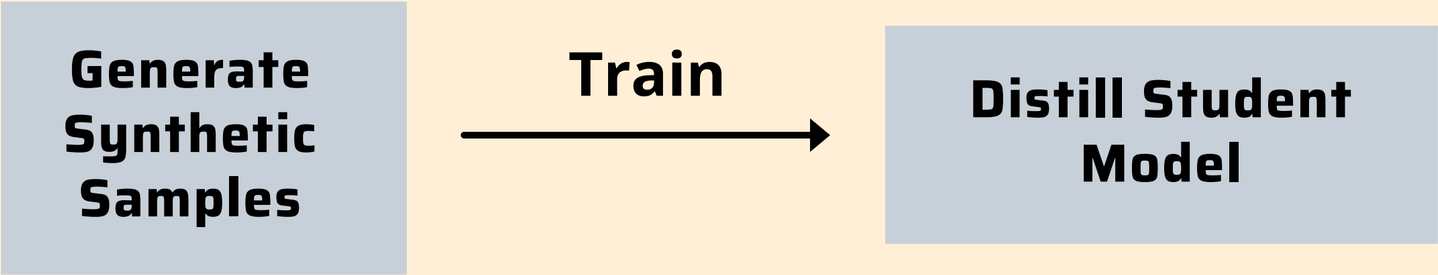
- The complexity of action recognition and the number of classes
- Lack of relevant literature on video classification model extraction
- Scale of datasets being used

Additional Challenges in Black Box Extraction

- No access to any dataset
- Lack of pre-trained Video GANS

Black Box Model Extraction

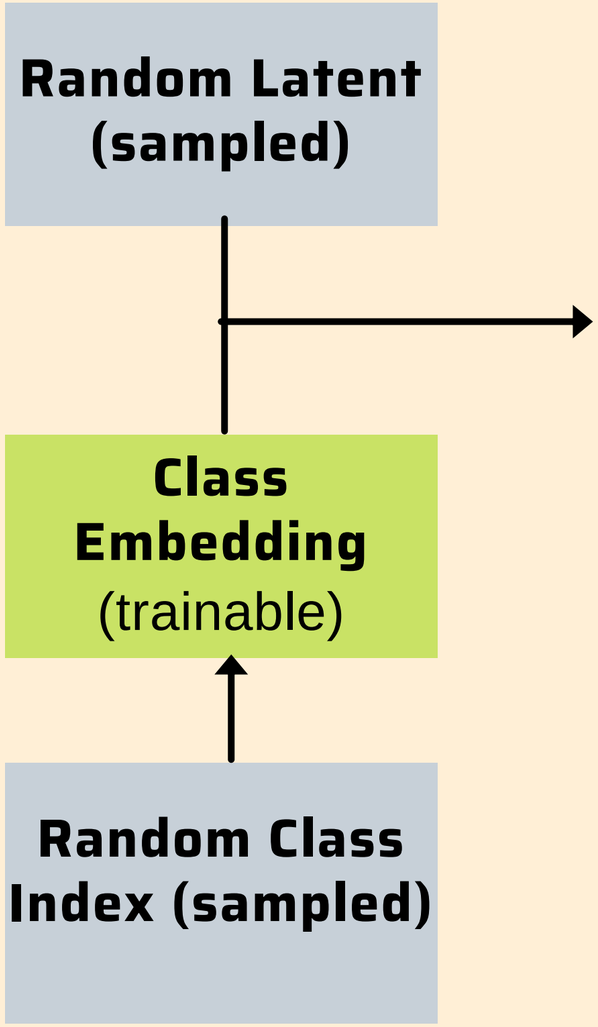
BLACK BOX APPROACH



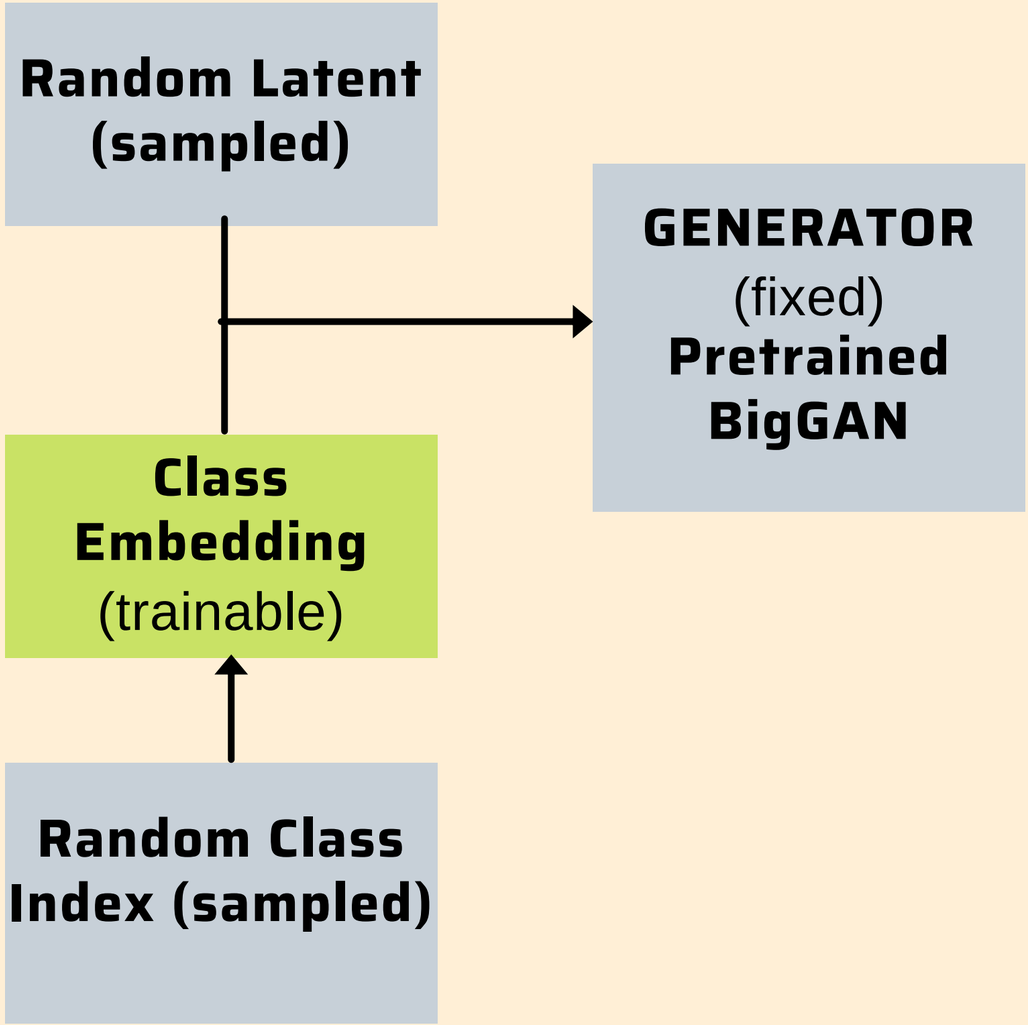
SYNTHETIC DATA GENERATION

- Due to the absence of class-conditioned Video GANs, we re-purpose a pre-trained BigGAN (ImageNet) to generate fake samples.
- Learn embeddings for each class index predicted by the teacher.
- Freeze generator, teacher model and finetune to minimise loss between teacher predicted class index and sampled class index.
- Finetune for a few steps until predicted class confidence > 90%

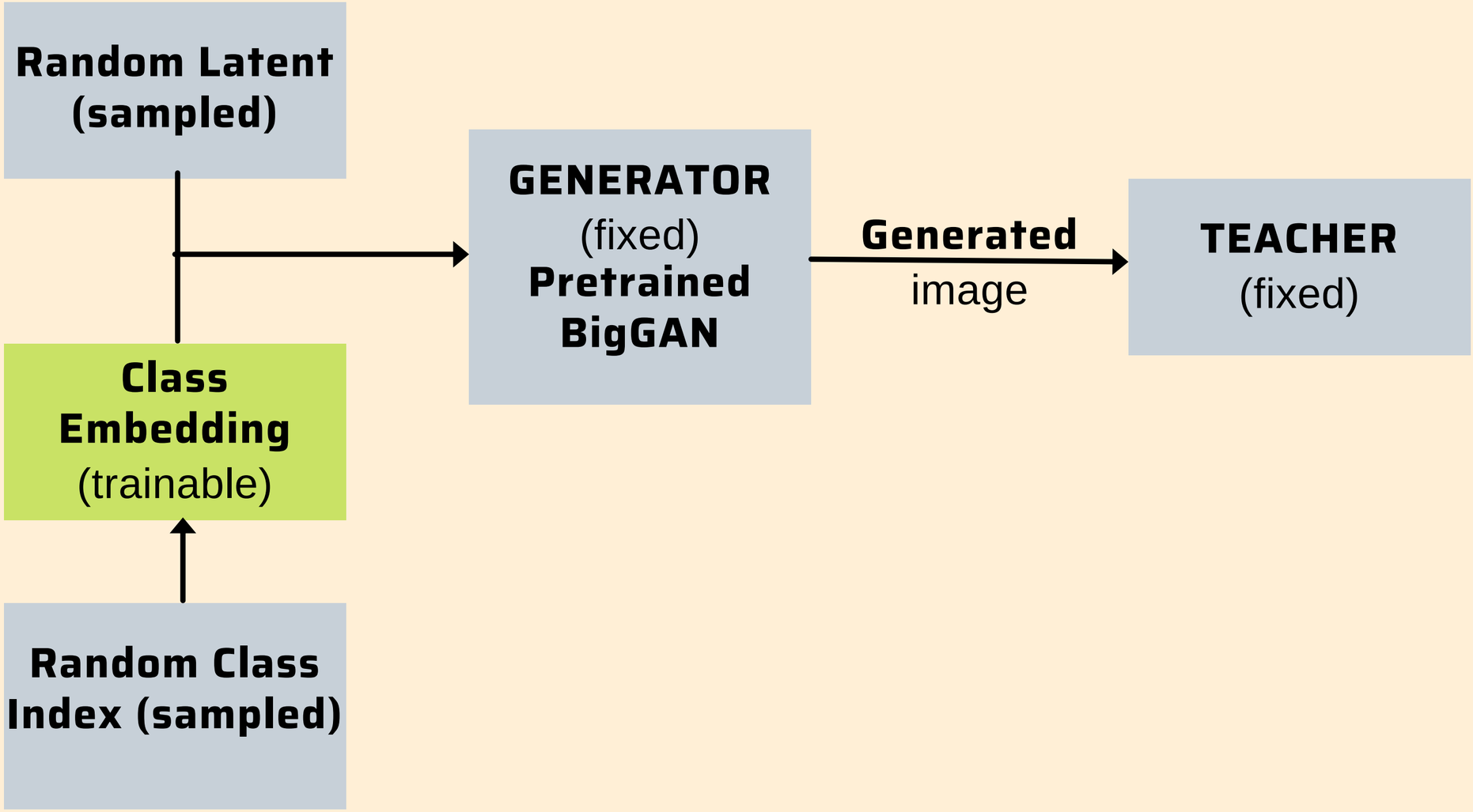
SYNTHETIC DATA GENERATION



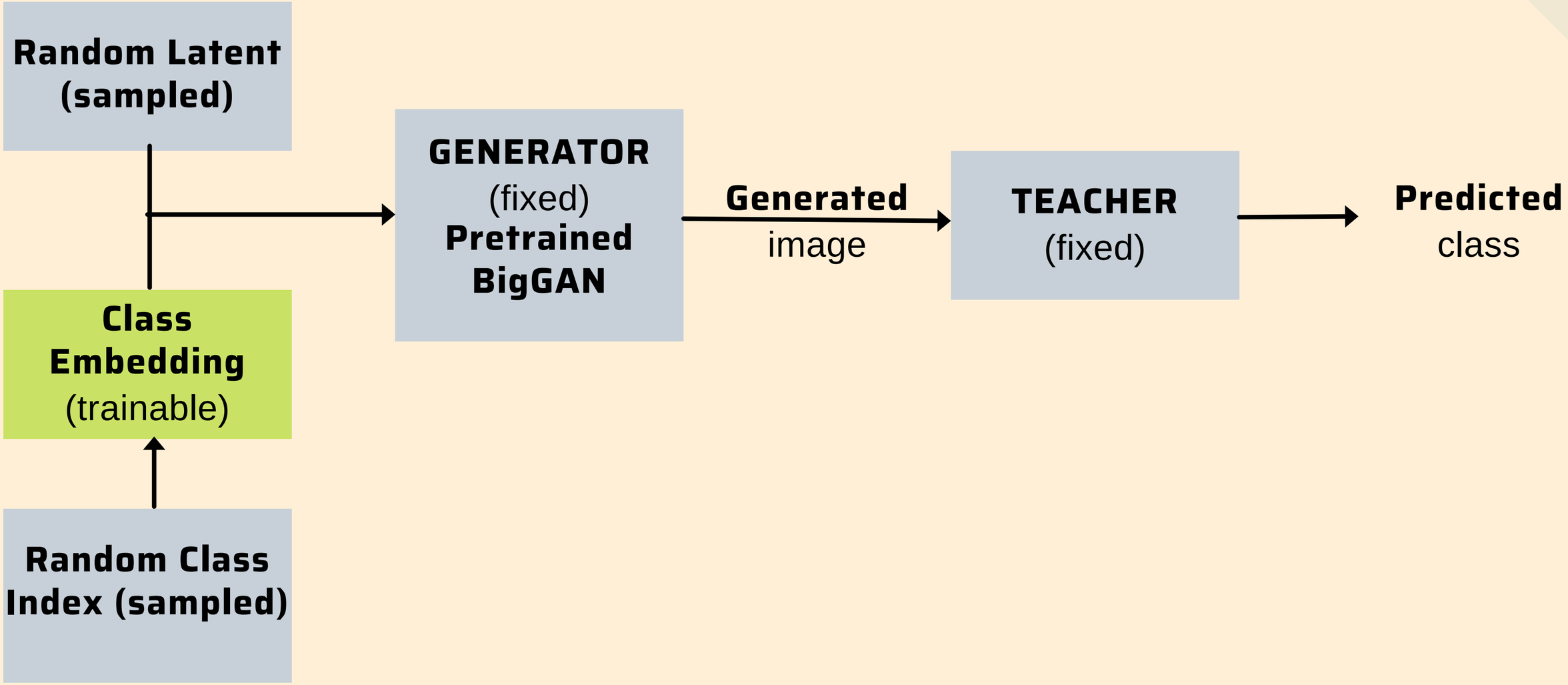
SYNTHETIC DATA GENERATION



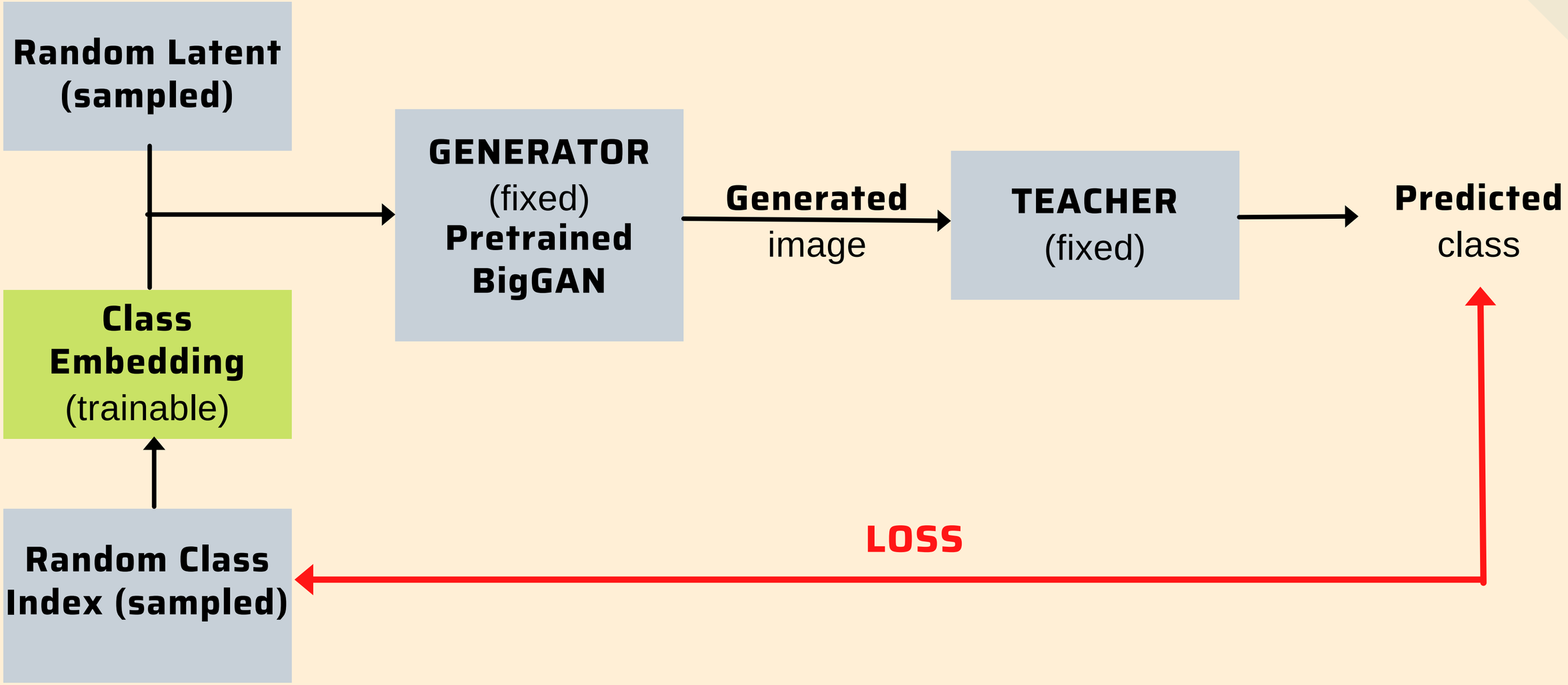
SYNTHETIC DATA GENERATION



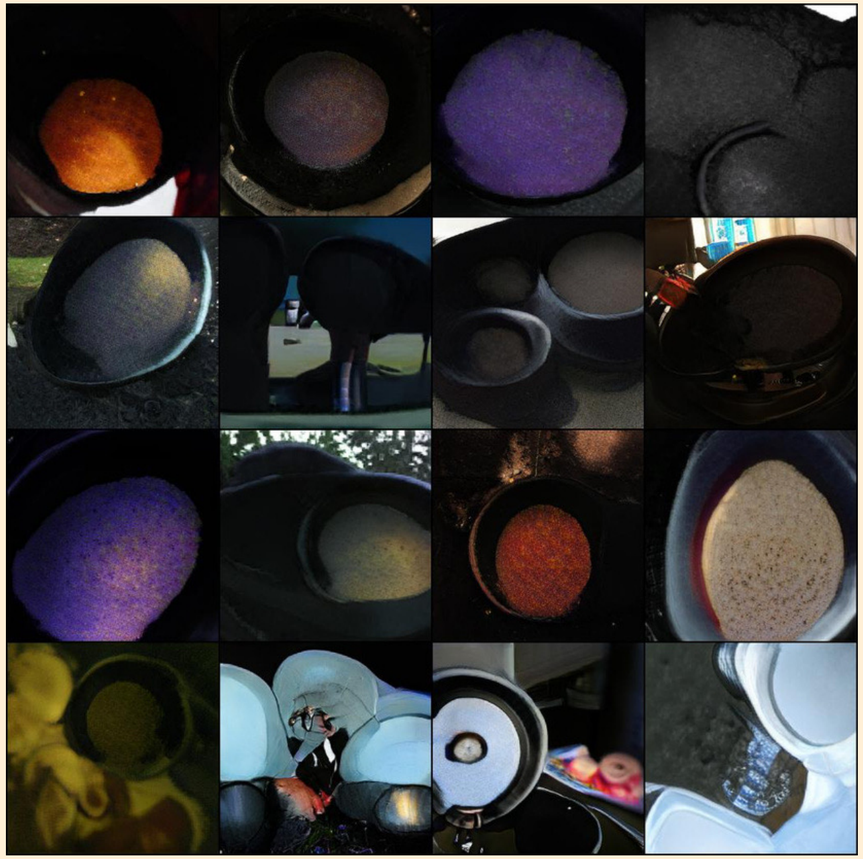
SYNTHETIC DATA GENERATION



SYNTHETIC DATA GENERATION

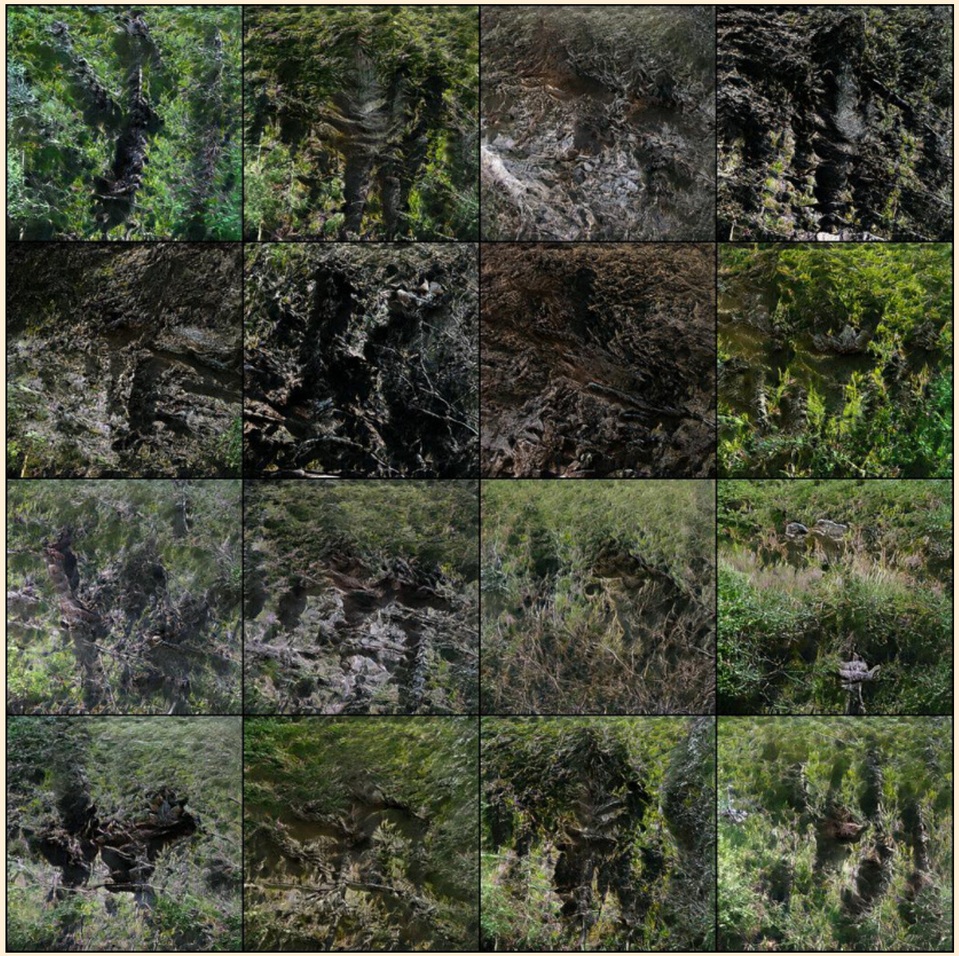


SYNTHETIC DATA - EXAMPLE



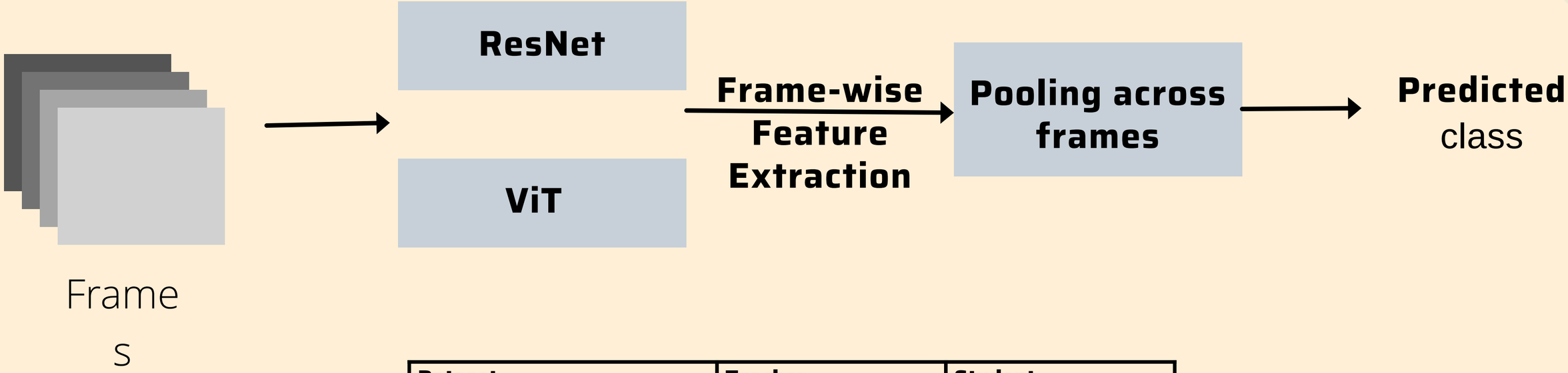
- The image corresponds to samples generated for class index 1 or "air drumming"
- Although the images are not visually appealing, they fool the teacher classifier with high confidence.
- Thus the generated samples belong to the training distribution of the teacher network.

SYNTHETIC DATA - EXAMPLE



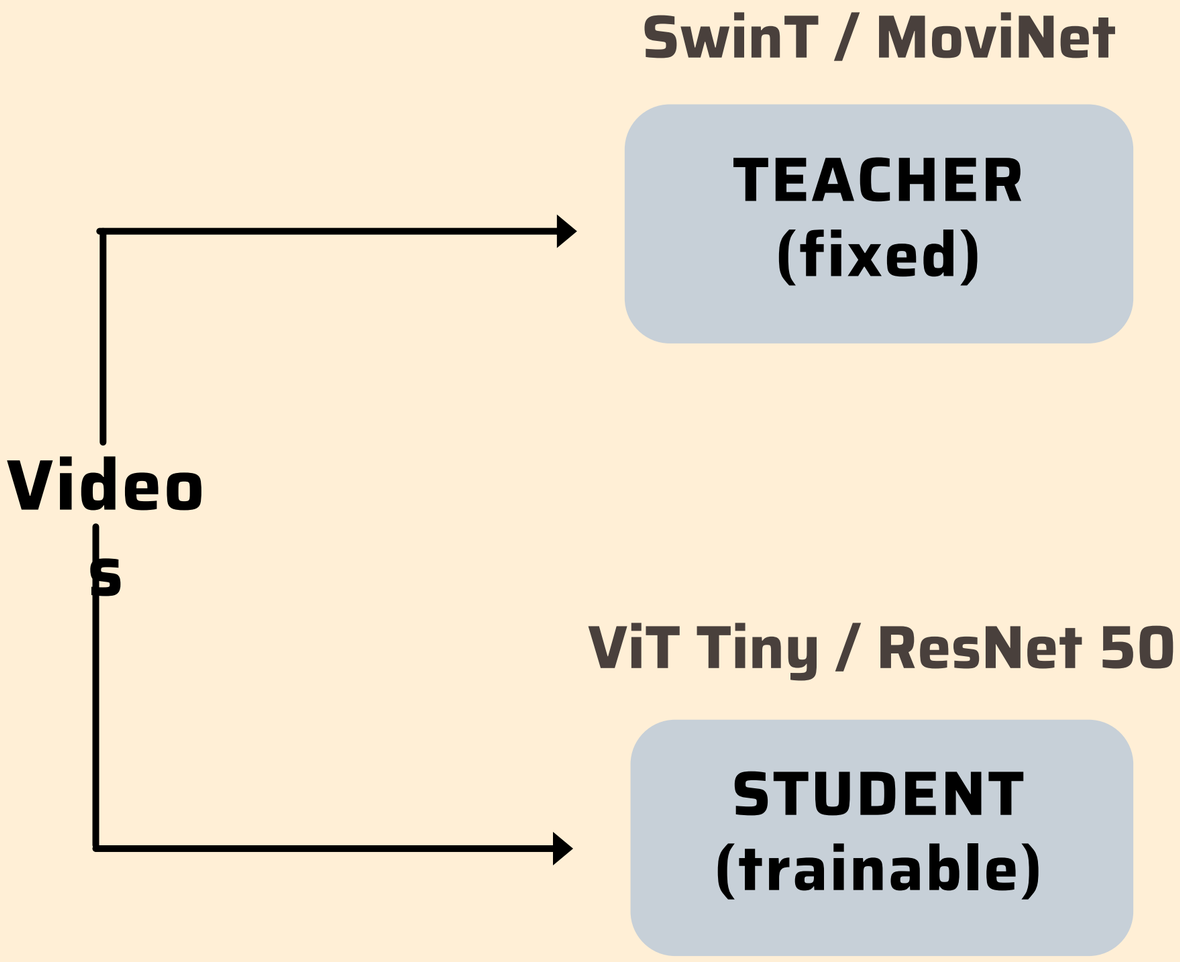
class index 0: abseiling

STUDENT MODEL

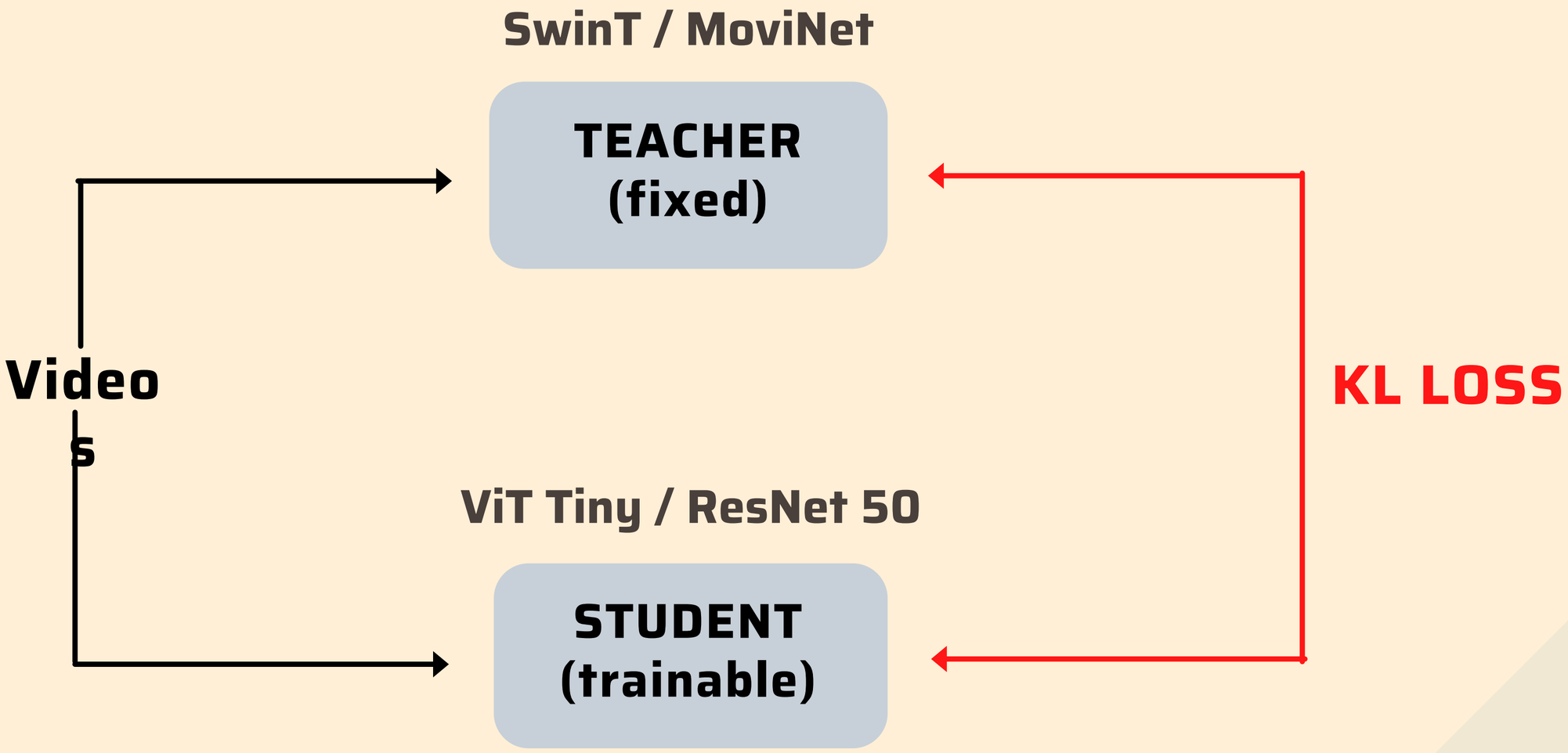


Dataset	Teacher	Student
Kinetics 600	MoviNet	ResNet-50
Kinetics 400	Swin-T	VIT-T

KNOWLEDGE DISTILLATION



KNOWLEDGE DISTILLATION

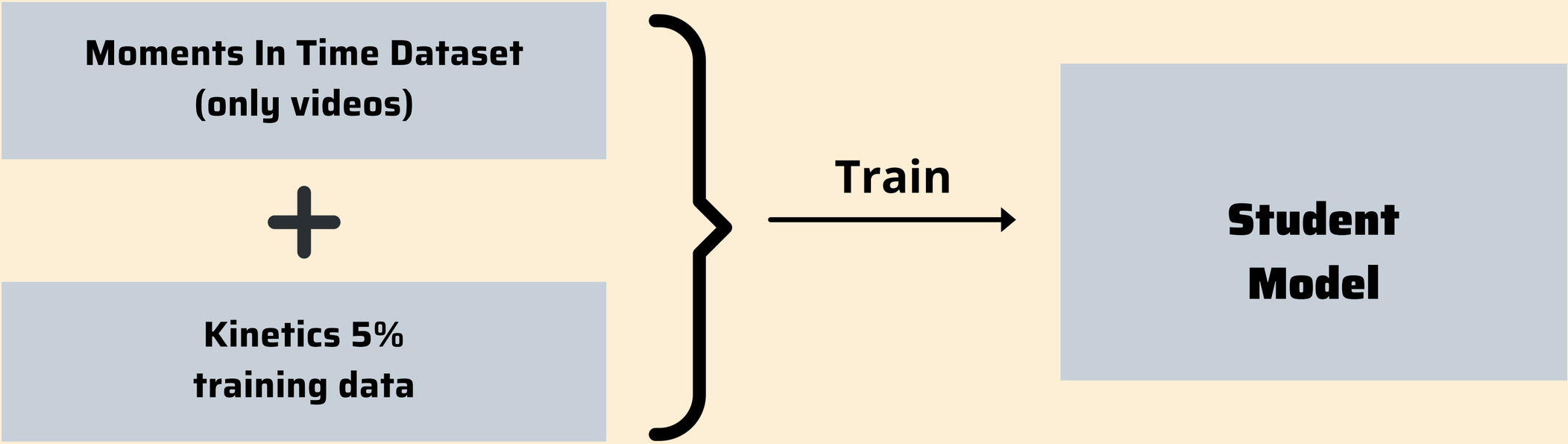


Black Box Accuracies

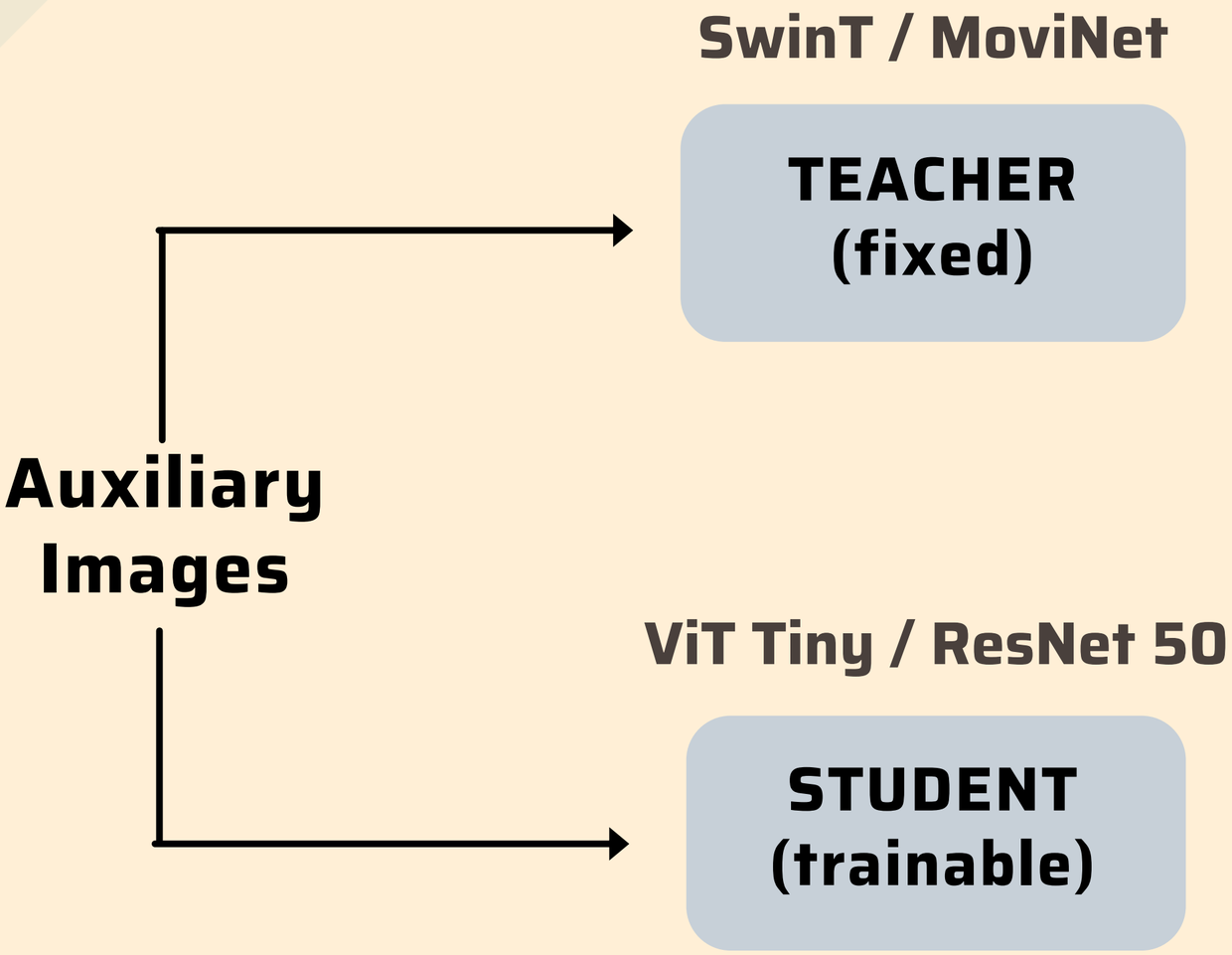
	VideoSwin **Transformer**	**Movinet**
Distill	**7.8%**	**12.43%**

Grey Box Model Extraction

GREY BOX APPROACH



KNOWLEDGE DISTILLATION



Auxillary Dataset KD

Irrelevant dataset: Moments in Time

- Number of classes: 305
- **Minimal** overlap with Kinetics 400 / Kinetics 600 Dataset

FINE TUNING ON 5% TRAINING DATA

Lottery Ticket Hypothesis, Why?

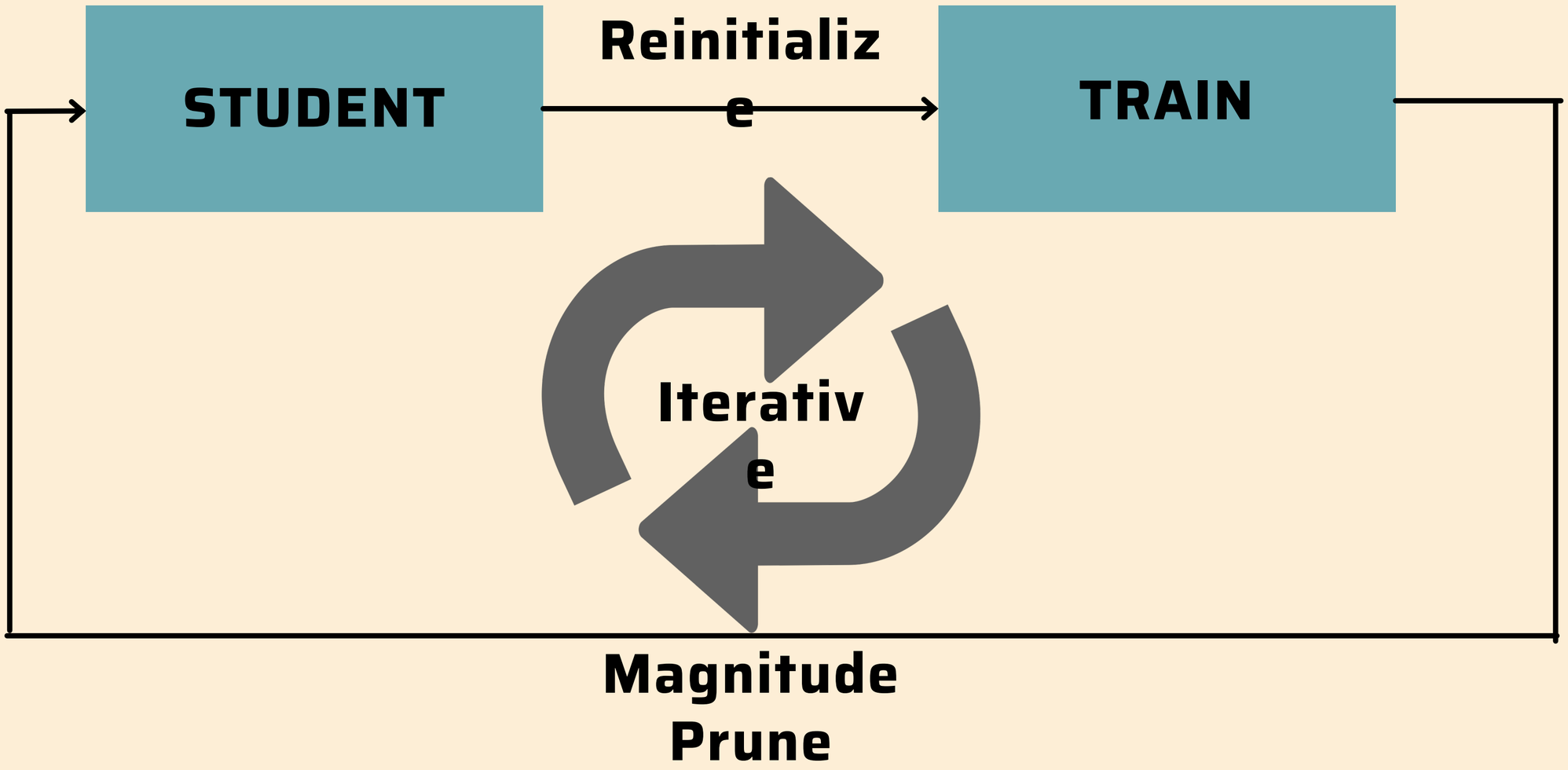
Every network contains a sparse network which can potentially outperform the dense network

- Sparsity is a regularization to avoid over-fitting.
- Induces inductive bias specific to fine-tuned task, hence improved performance.

Iterative Magnitude Pruning

- Initialize network
- Train for a few epochs
- Prune least magnitude weights
- Re-initialize and perform steps 2-3 till required sparsity.

FINE TUNING ON 5% TRAINING DATA



Grey Box Accuracies

	VideoSwin **Transformer**	**Movinet**
Knowledge **Distillation**	**8.3%**	**13.8%**
Fine-Tune 5% Kinetics **(Dense)**	**12.6%**	**22.3%**
Fine-Tune 5% Kinetics **(Winning Ticket)**	**17.3%**	**33.5%**

Thanks

Team 6

!