Research Seminar Coding Task

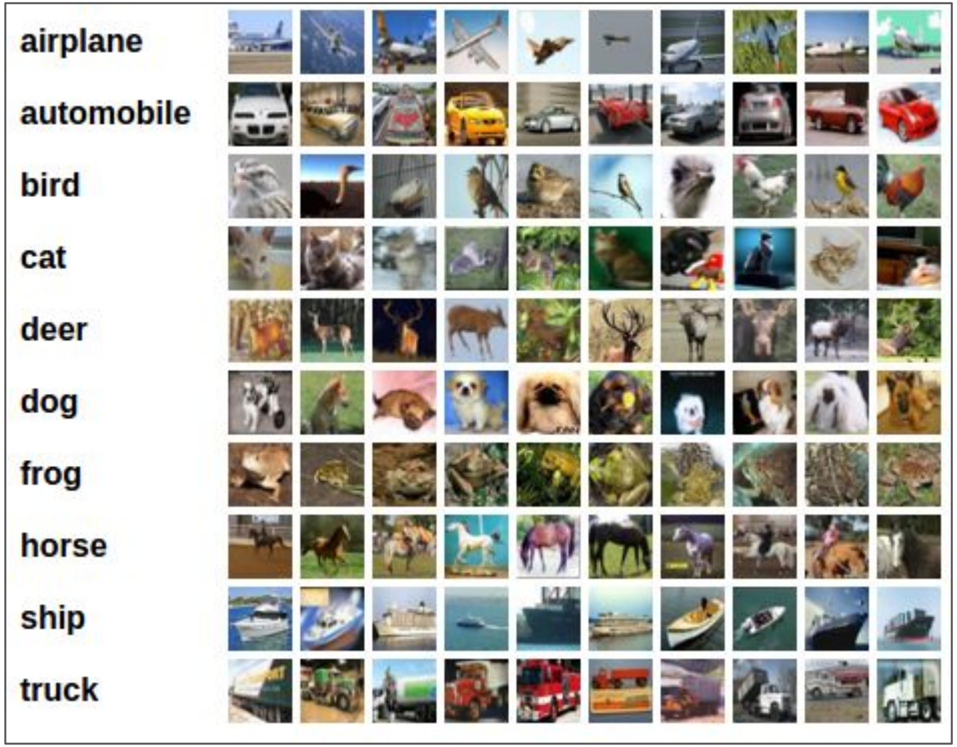
Y Data Program

Asaf Dahan and Omer Dodi

March 2022

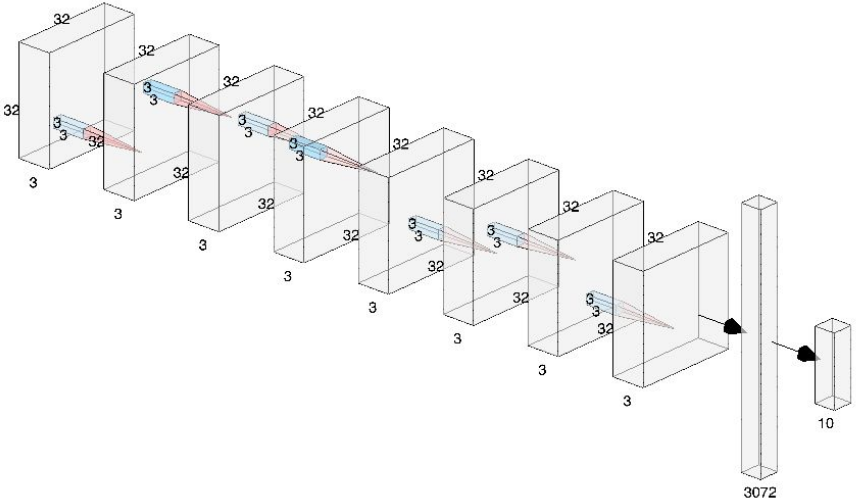
Dataset

- CIFAR-10
- 60K Images
  - 50K train dataset
  - 10K validation dataset
- 32x32 Pixels, 3 channels
- 10 Classes
  - Balanced
- Split datasets by classes
  - A with 5 classes
  - B with 5 classes
  - Randomly splitted to 5 groups



Initial Architecture

- 8x Convolutions
  - 3 filters (3x3)
  - Padding + stride of 1
  - ReLu activation
- Flatten to 3072 -> 10

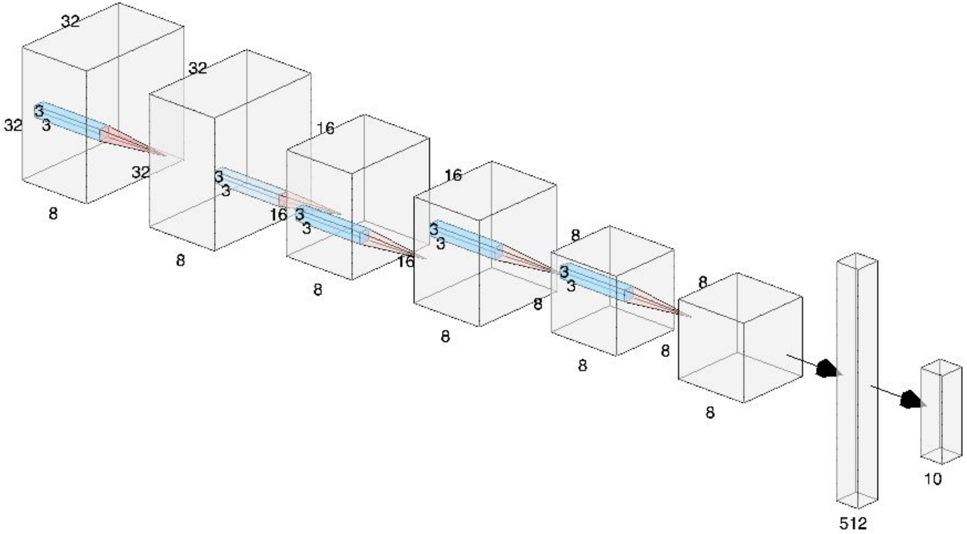


The problem:

- We got low baseline accuracy

Final Architecture

- 3x of:
  - 2x Convolutions
    - 8 filters (3x3)
    - Padding + stride of 1
    - Batch normalization
    - ReLu activation
  - Max pooling 2d (2x2)
- Flatten to 512 -> 10

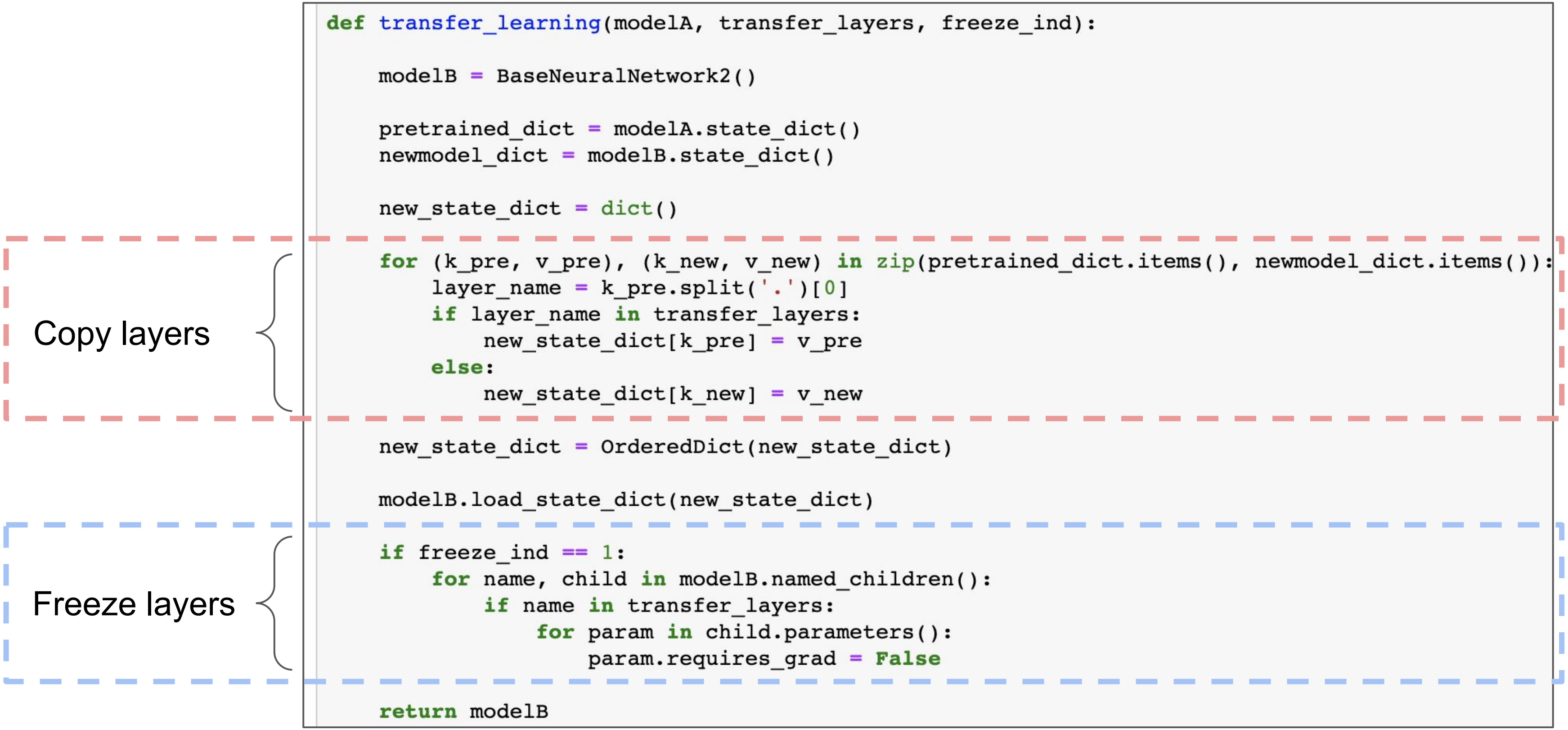


- Baseline accuracy (10 epochs) - 0.689

Workflow

- Load data
- Split data (A/B)
    - 5 splits
- Build our own NN architecture
- Train base models
    - 10 models
- **Create transfer learning function**
- Train experimental models
    - B3B, B3B+ (“Selffer”), A3B, A3B+ (“Transfer”)
    - ~160 models (2 freeze ind * 8 transfer layers * 10 base models)
- Draw graphs
- Get insights

Experiments

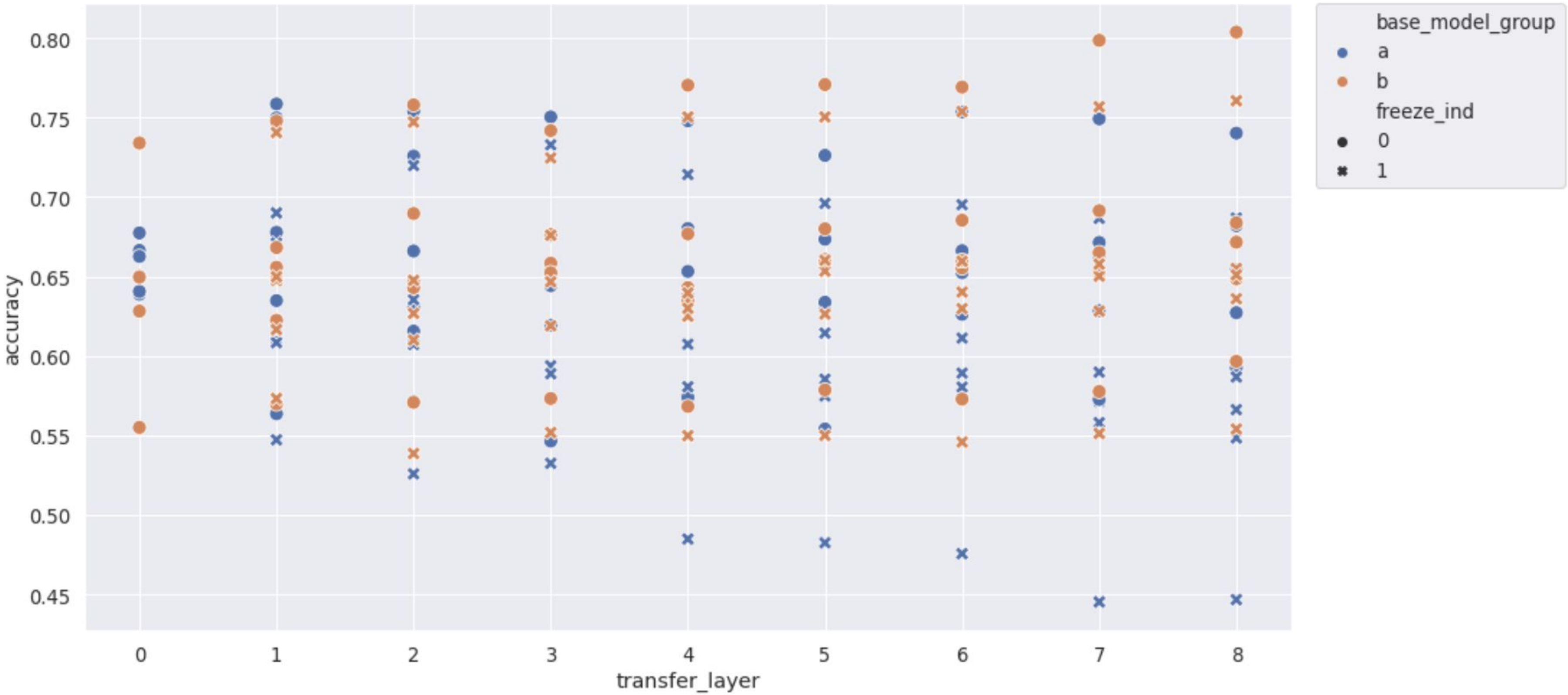


Copy layers

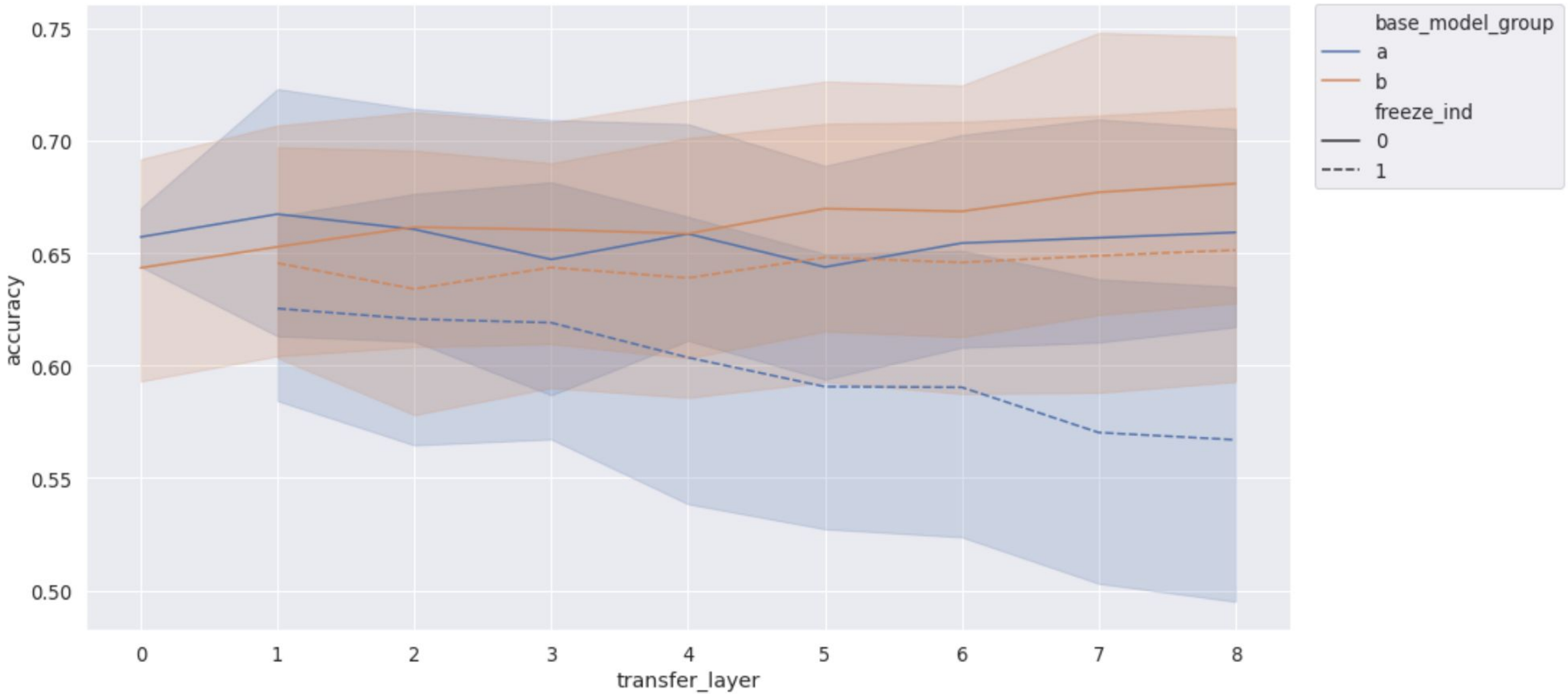
Freeze layers

```python
def transfer_learning(modelA, transfer_layers, freeze_ind):

    modelB = BaseNeuralNetwork2()

    pretrained_dict = modelA.state_dict()
    newmodel_dict = modelB.state_dict()

    new_state_dict = dict()

    for (k_pre, v_pre), (k_new, v_new) in zip(pretrained_dict.items(), newmodel_dict.items()):
        layer_name = k_pre.split('.')[0]
        if layer_name in transfer_layers:
            new_state_dict[k_pre] = v_pre
        else:
            new_state_dict[k_new] = v_new

    new_state_dict = OrderedDict(new_state_dict)

    modelB.load_state_dict(new_state_dict)

    if freeze_ind == 1:
        for name, child in modelB.named_children():
            if name in transfer_layers:
                for param in child.parameters():
                    param.requires_grad = False

    return modelB
```

Results


accuracy
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0
1
2
3
4
5
6
7
8
transfer_layer
base_model_group
a
b
freeze_ind
0
1

Results



0.75
0.70
0.65
0.60
0.55
0.50
accuracy
0
1
2
3
4
5
6
7
8
transfer_layer
base_model_group
a
b
freeze_ind
0
1

Conclusions

- Finetune yield better results

As for conclusions from the original paper

✗ Transfer + finetune improve generalization

✗ Performance drops due to fragile co-adaptation

✓ Performance drops due to representation specificity

Possible explanations

- Due to small dataset - low number of samples and number of classes
- Due to low number of epochs - 3