Ttansfer Learning

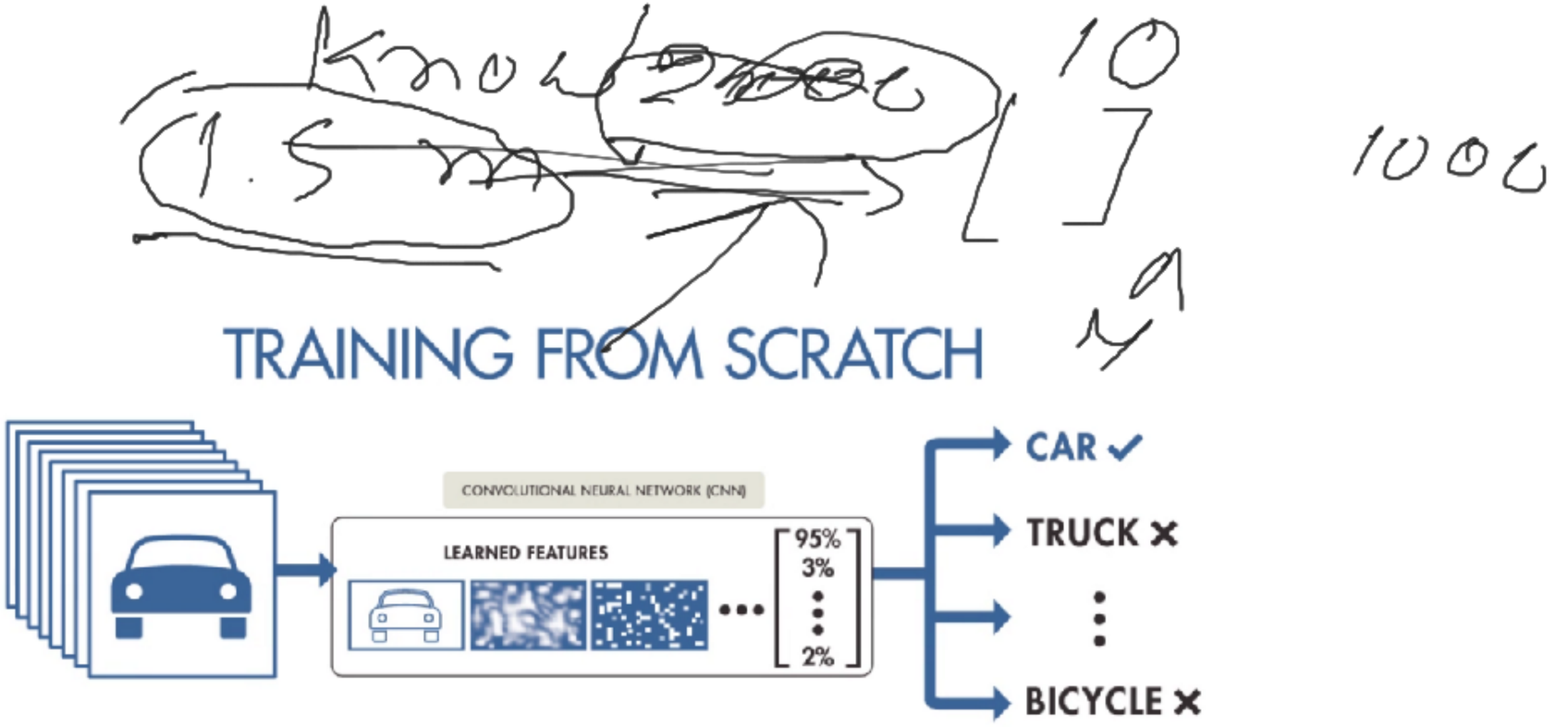
What is it?

How does it work?

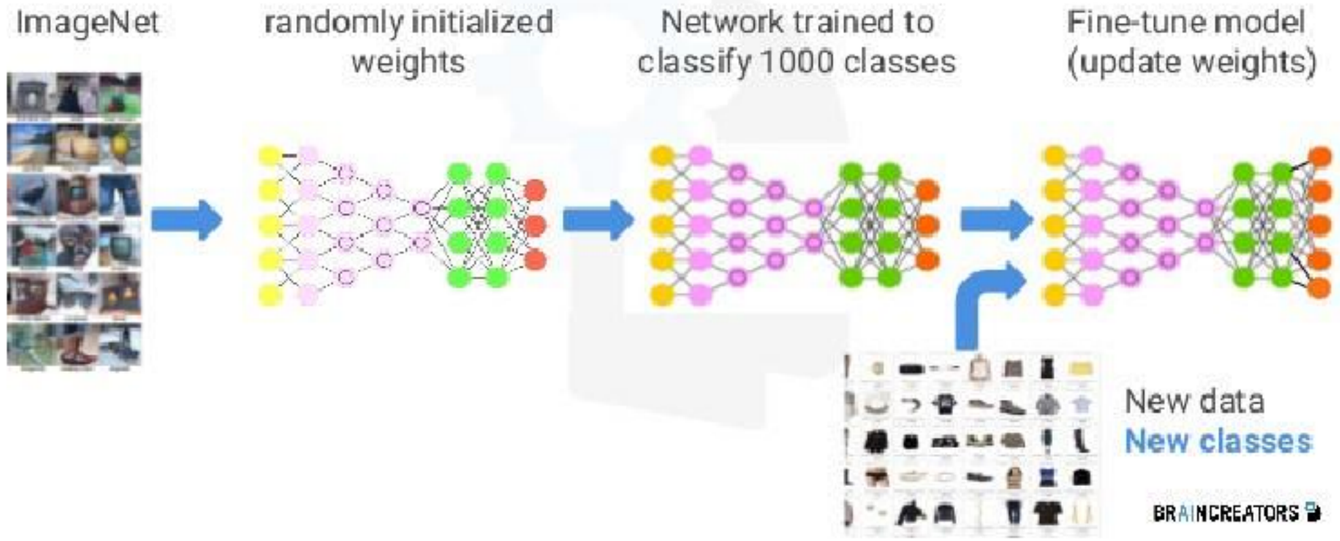
Why it is used?

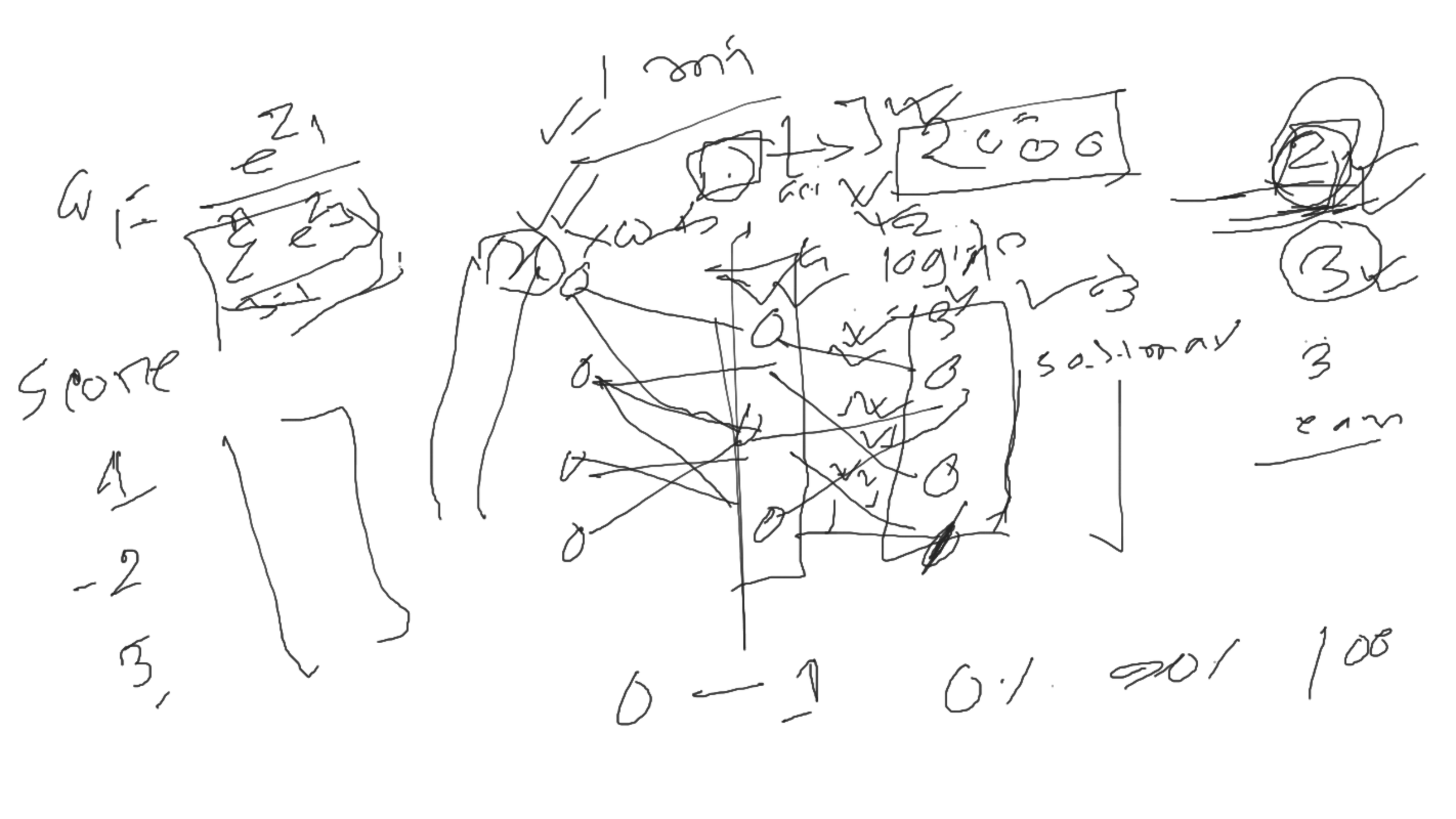
When should you use it?

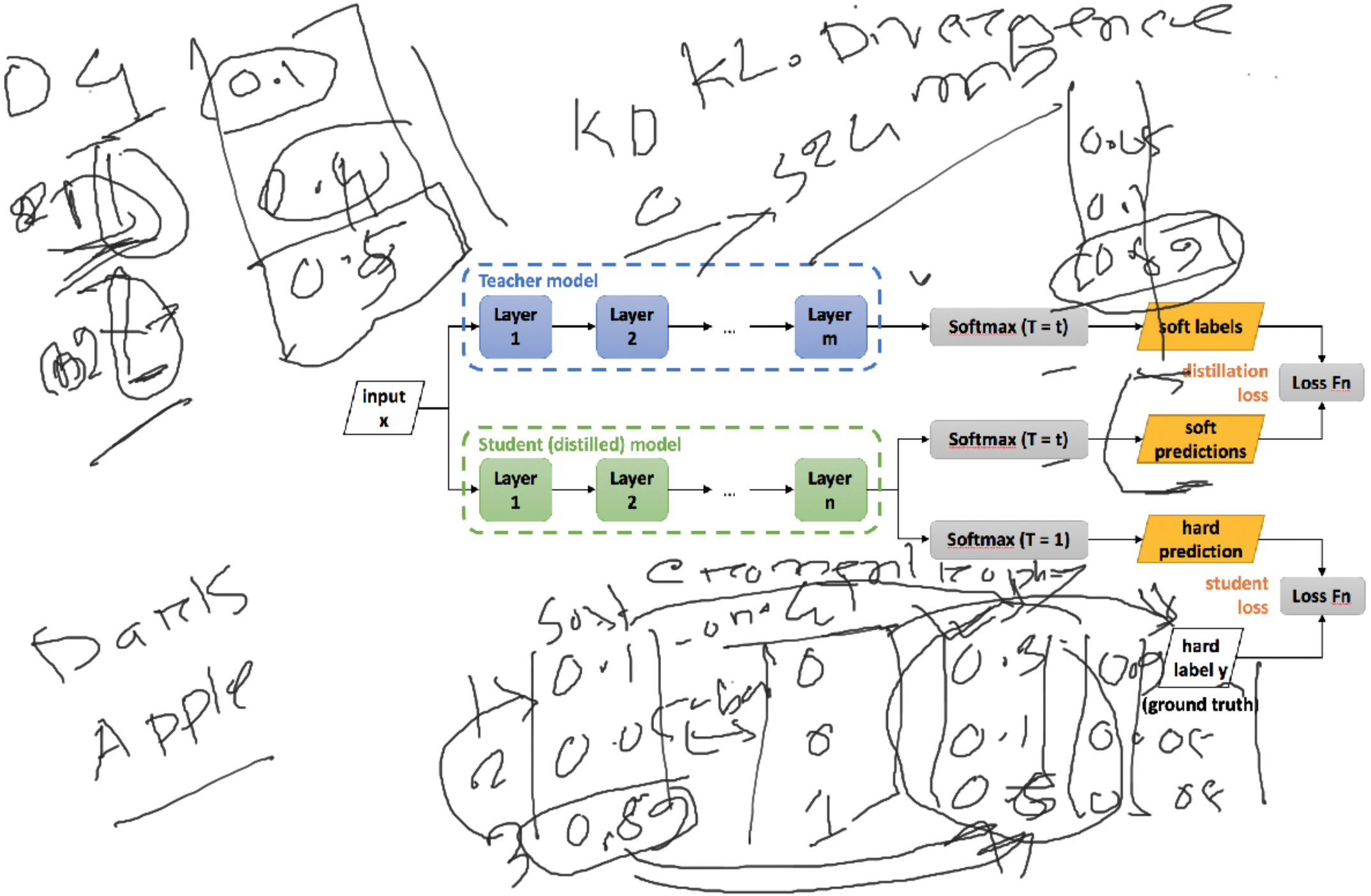
Approaches to transfer learning: Training a model to reuse it; Using a pre-trained model; Feature extraction


TRAINING FROM SCRATCH
CONVOLUTIONAL NEURAL NETWORK (CNN)
LEARNED FEATURES
95%
3%
2%
CAR ✓
TRUCK ✗
BICYCLE ✗
TRANSFER LEARNING
TRAINED ON CATS AND DOGS
FINE-TUNE NETWORK WEIGHTS
PRE-TRAINED CNN
NEW TASK
CAR ✓
TRUCK ✗
NEW DATA

Transfer Learning




softmax


Teacher model
Layer 1
Layer 2
...
Layer m
Softmax (T = t)
soft labels
distillation loss
Loss Fn
input x
Student (distilled) model
Layer 1
Layer 2
...
Layer n
Softmax (T = t)
soft predictions
Softmax (T = 1)
hard prediction
student loss
Loss Fn
hard label y
(ground truth)

524mb

GT

64mb

0
0
1

Device 30 →

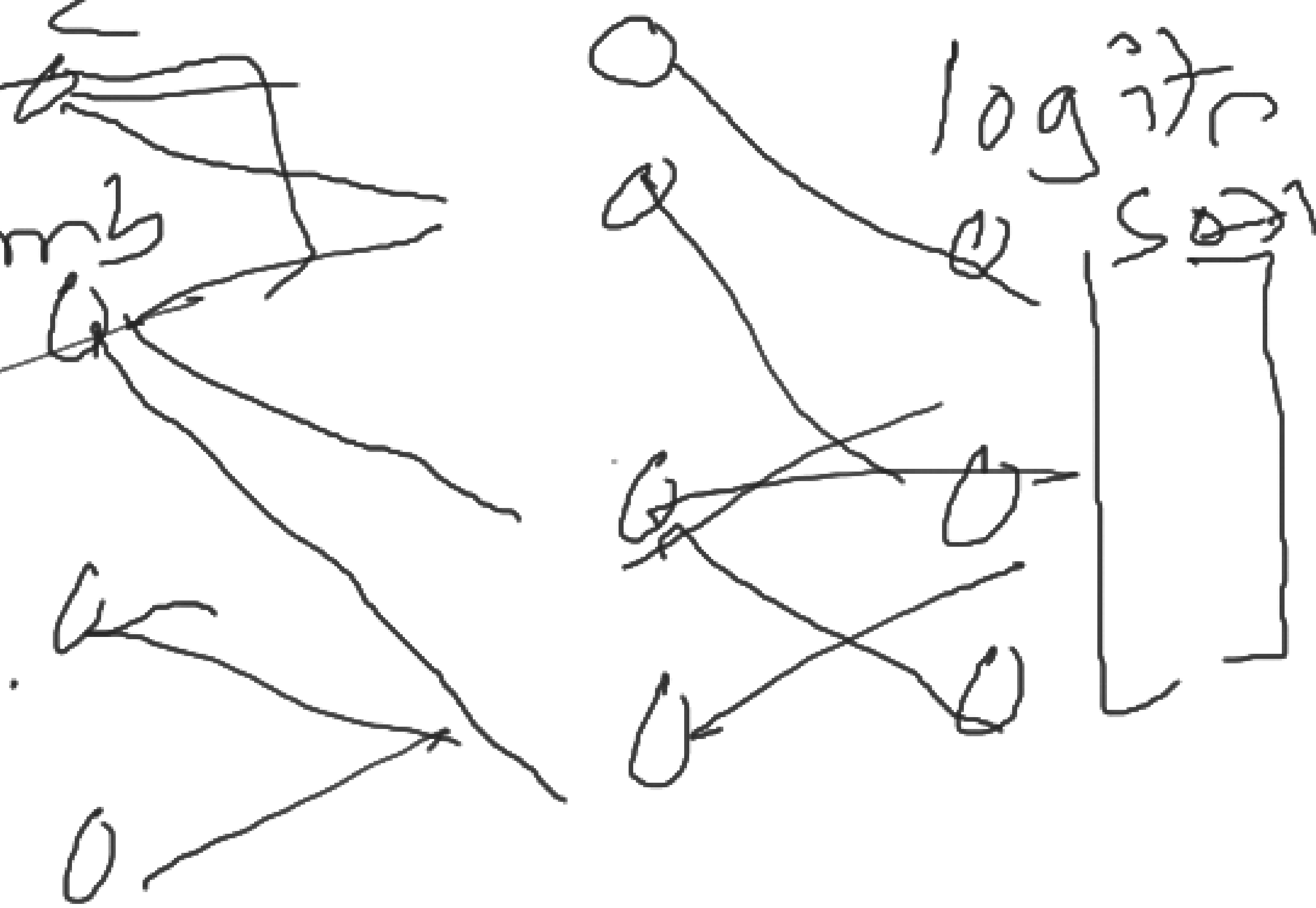
584 | 584 mb

Sot

0.1
0.05
0.88

74%

64mb



30 class

31%

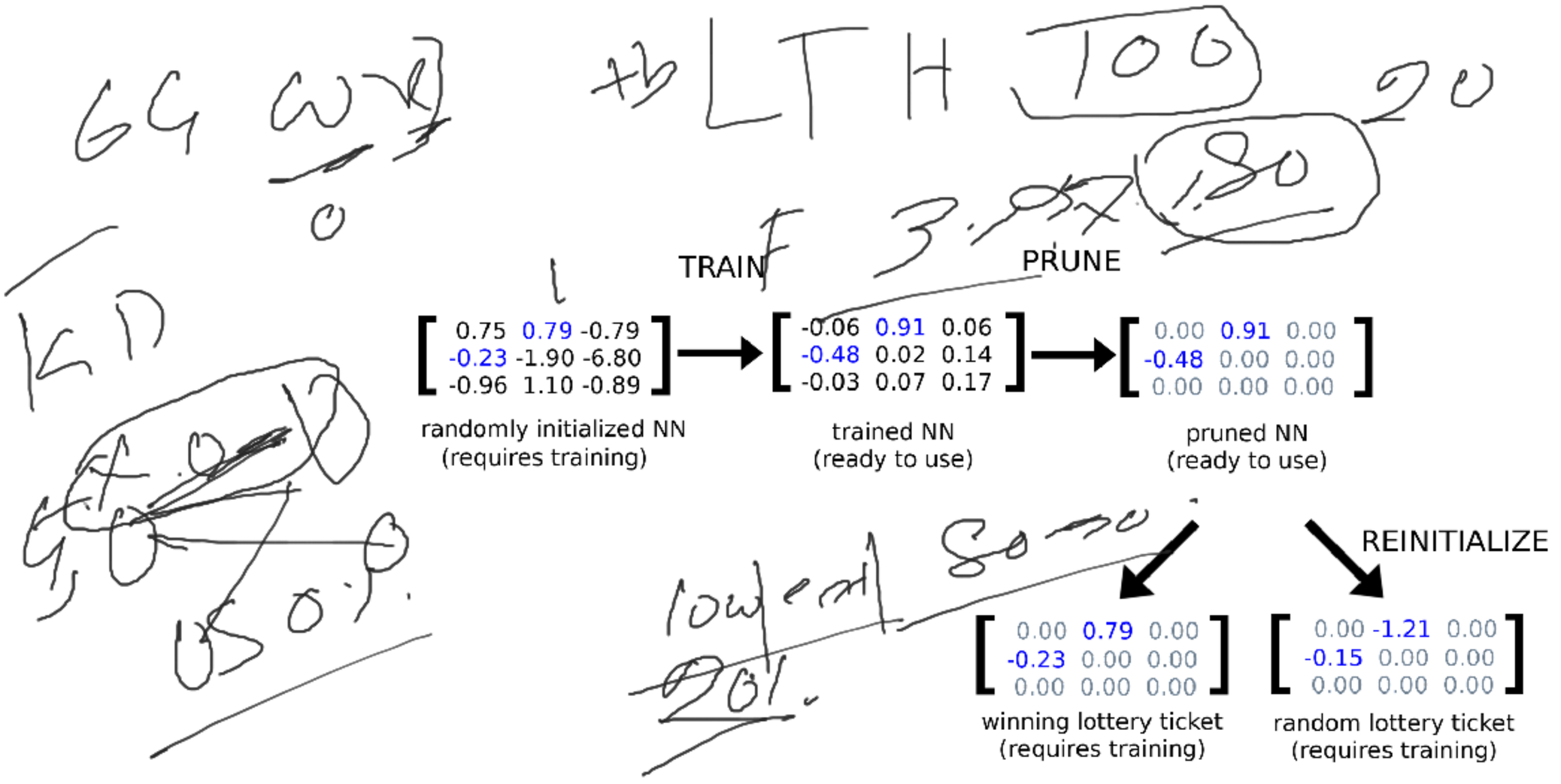
CE → (TP, Gr)

(SP, TP)

100% + CE + D L

$(1-\alpha) \times CE$ (0.7)

$(+) DL \times (-1 + alpha)$


TRAIN
PRUNE
0.75 0.79 -0.79
-0.23 -1.90 -6.80
-0.96 1.10 -0.89
randomly initialized NN
(requires training)
-0.06 0.91 0.06
-0.48 0.02 0.14
-0.03 0.07 0.17
trained NN
(ready to use)
0.00 0.91 0.00
-0.48 0.00 0.00
0.00 0.00 0.00
pruned NN
(ready to use)
REINITIALIZE
0.00 0.79 0.00
-0.23 0.00 0.00
0.00 0.00 0.00
winning lottery ticket
(requires training)
0.00 -1.21 0.00
-0.15 0.00 0.00
0.00 0.00 0.00
random lottery ticket
(requires training)

The Lottery Ticket Hypothesis (LTH)

Postulates existence of sparse trainable subnetworks which match original network accuracy.