ECS 289 Paper Presentation

“How transferable are features in deep neural networks?”

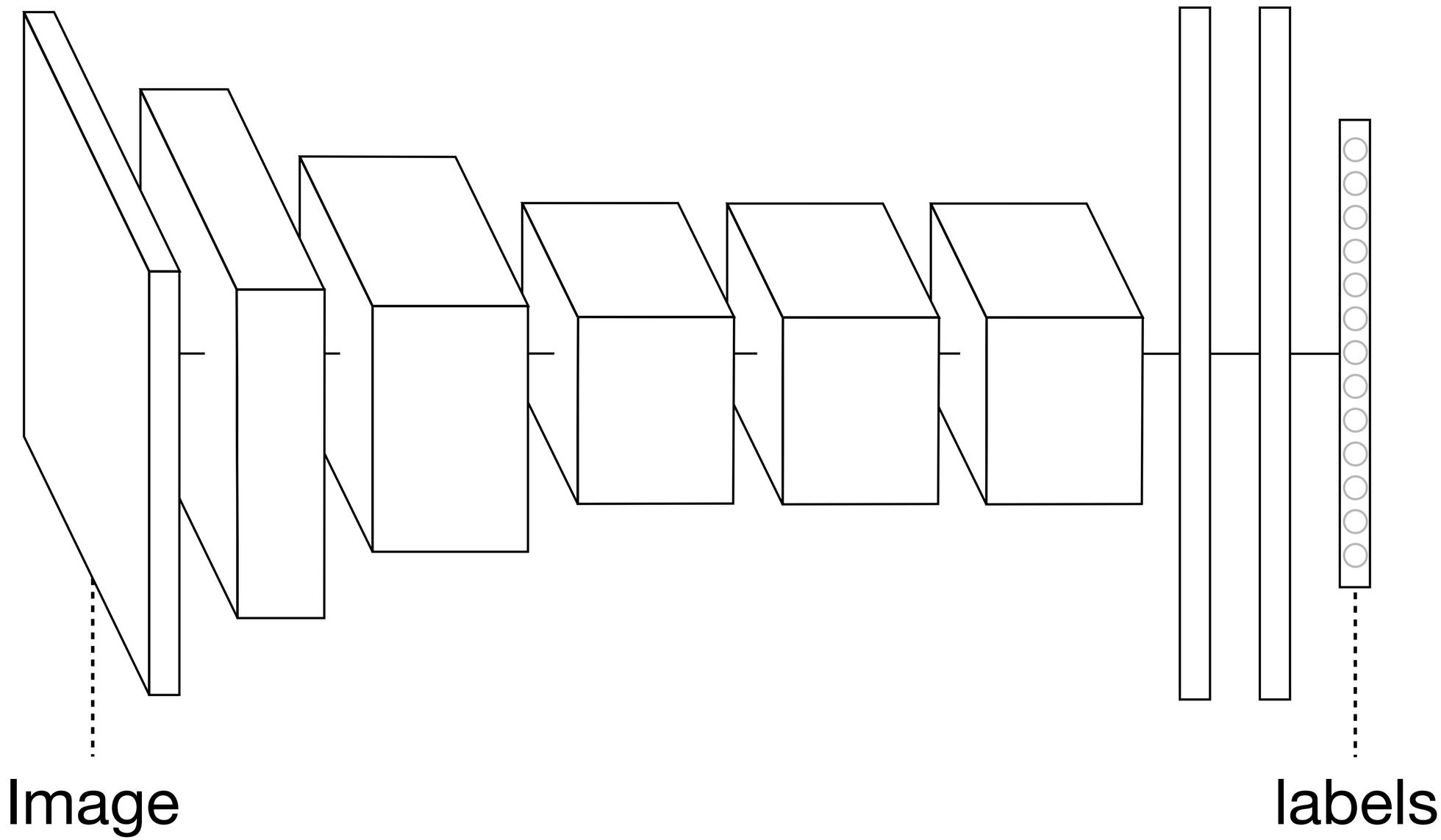
Written by Jason Yosinski, Jeff Clune, Yoshua Bengio, Hod Lipson

Present by Shubo Chen

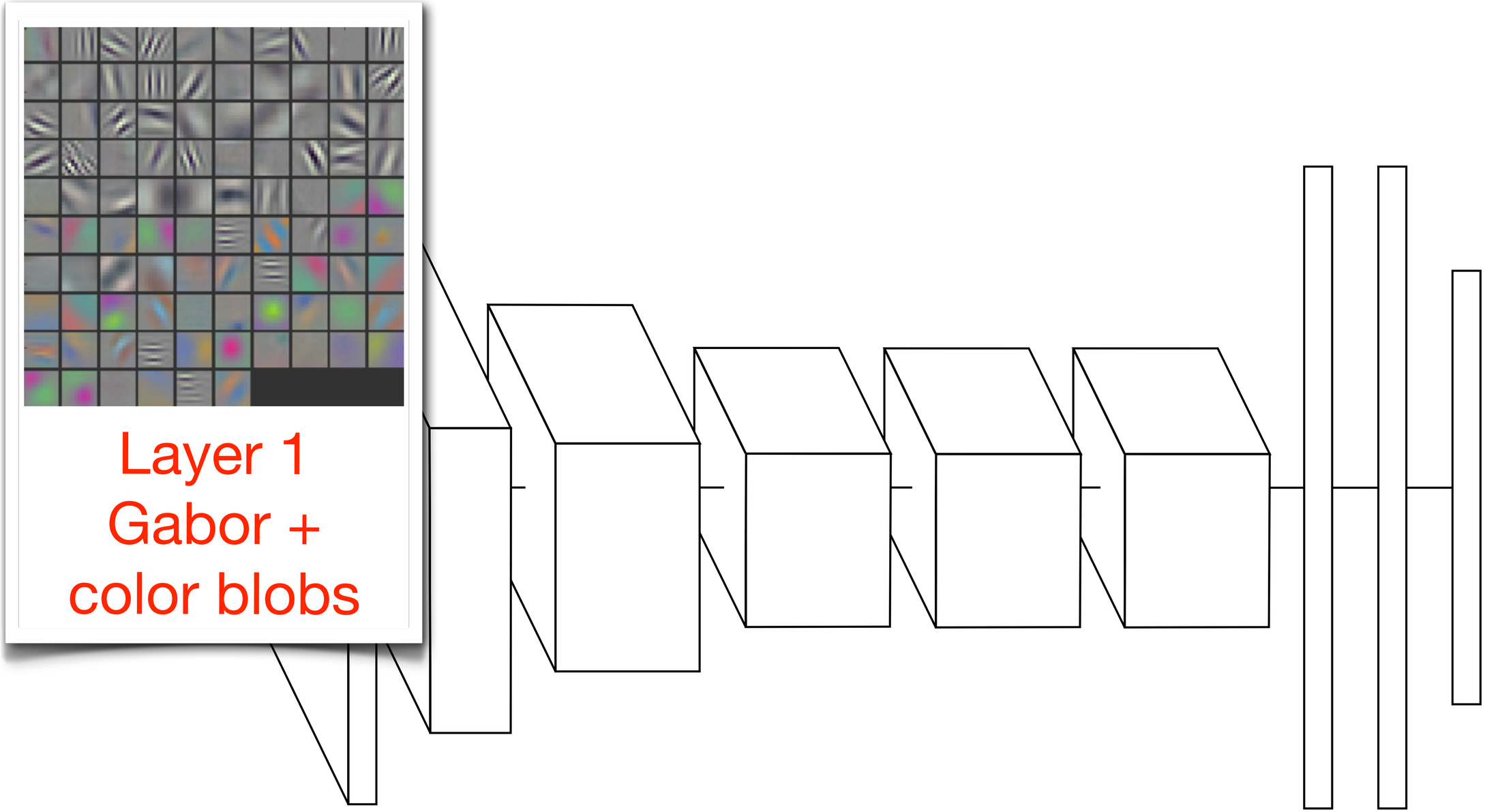
Date: /10/6/2016

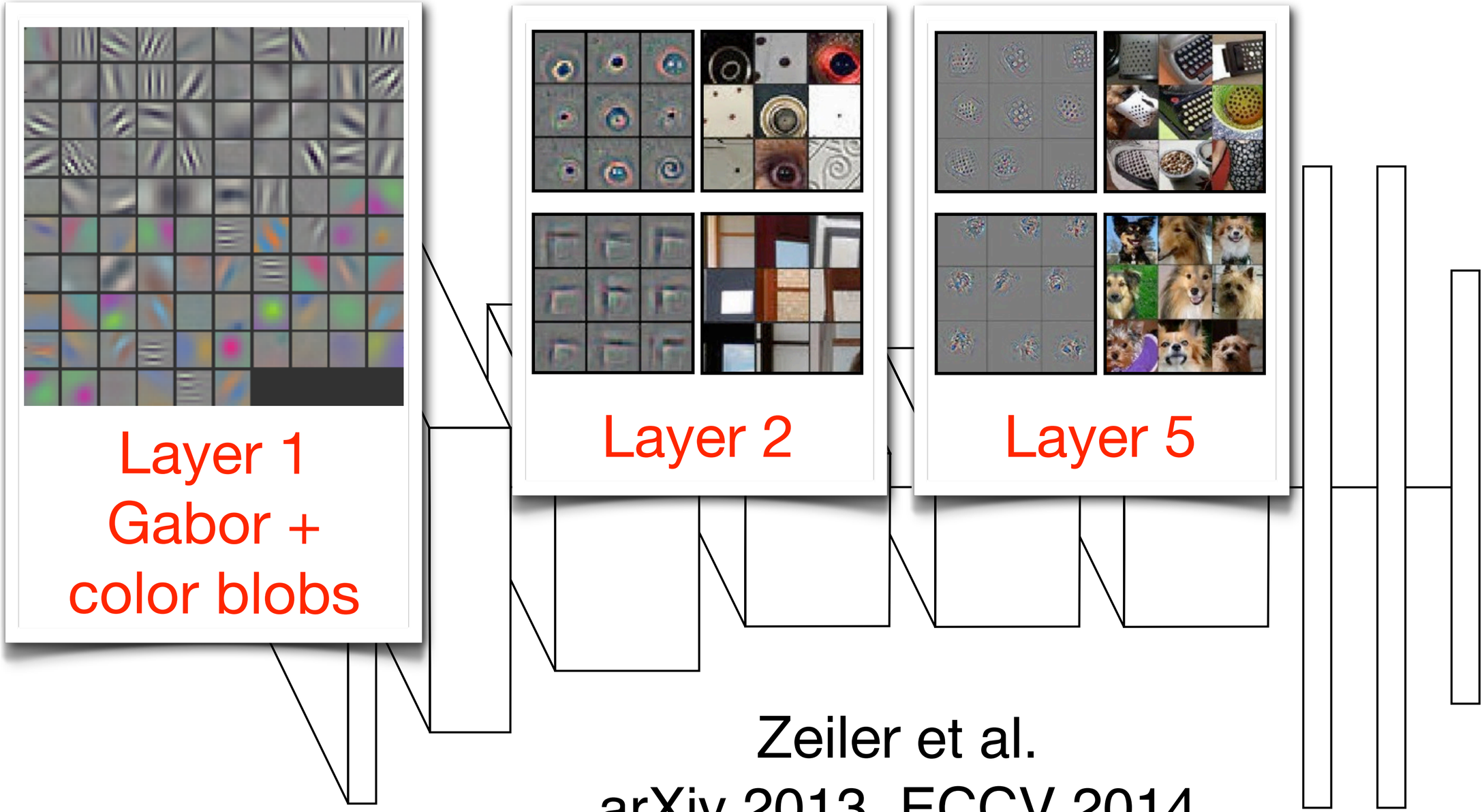
Slides originally made by Jason Yosinski and his collegues

AlexNet

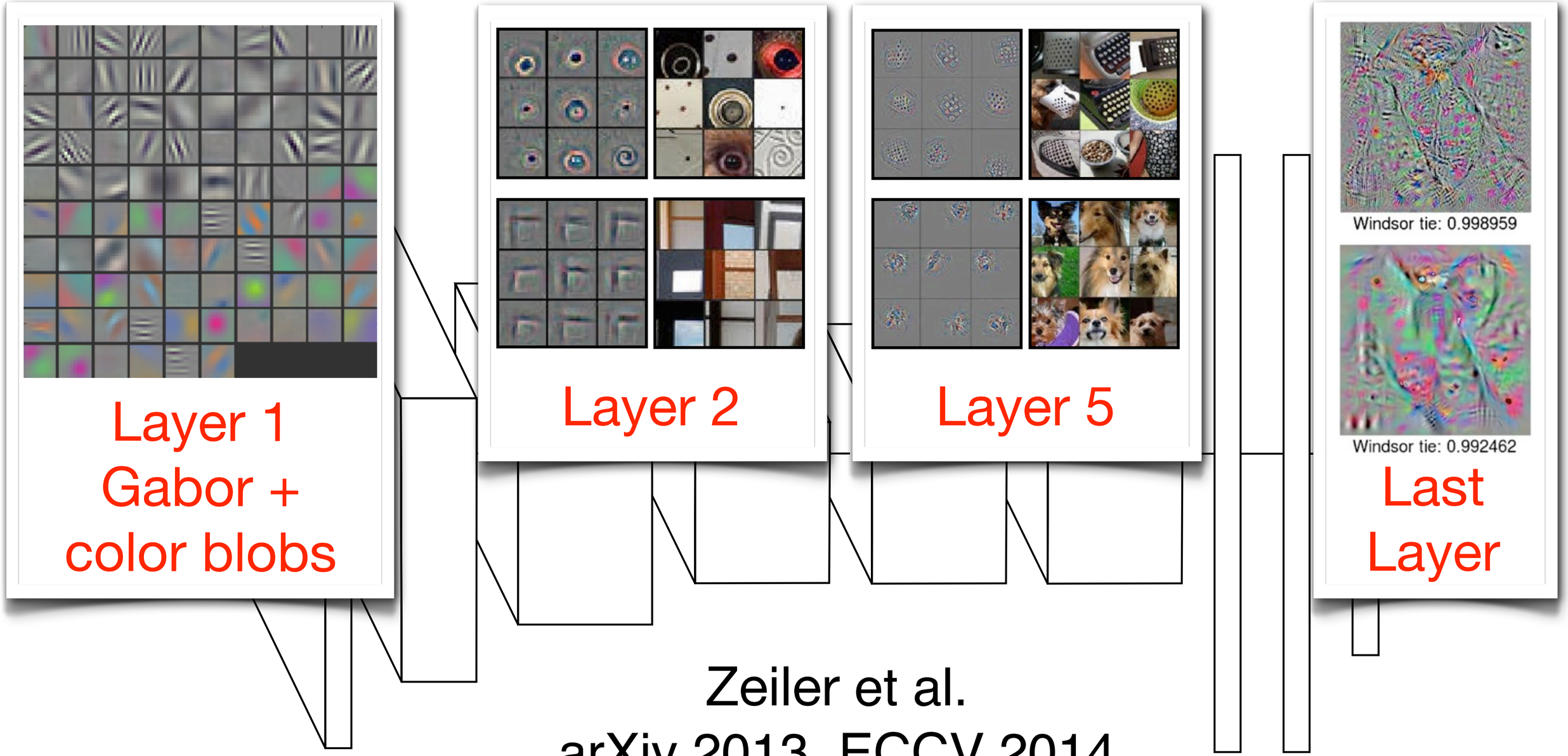


Krizhevsky, Sutskever, Hinton — NIPS 2012


Layer 1
Gabor +
color blobs

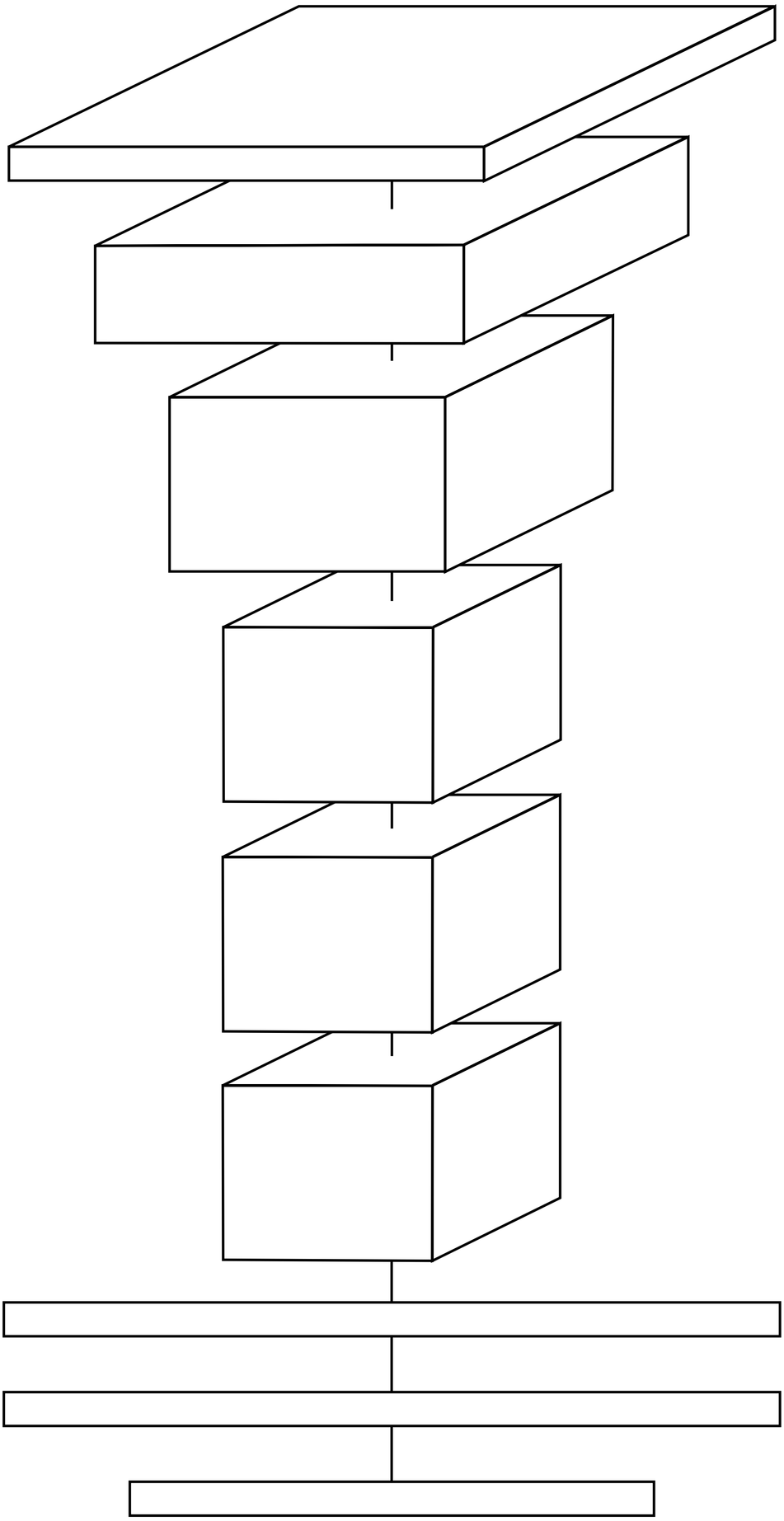


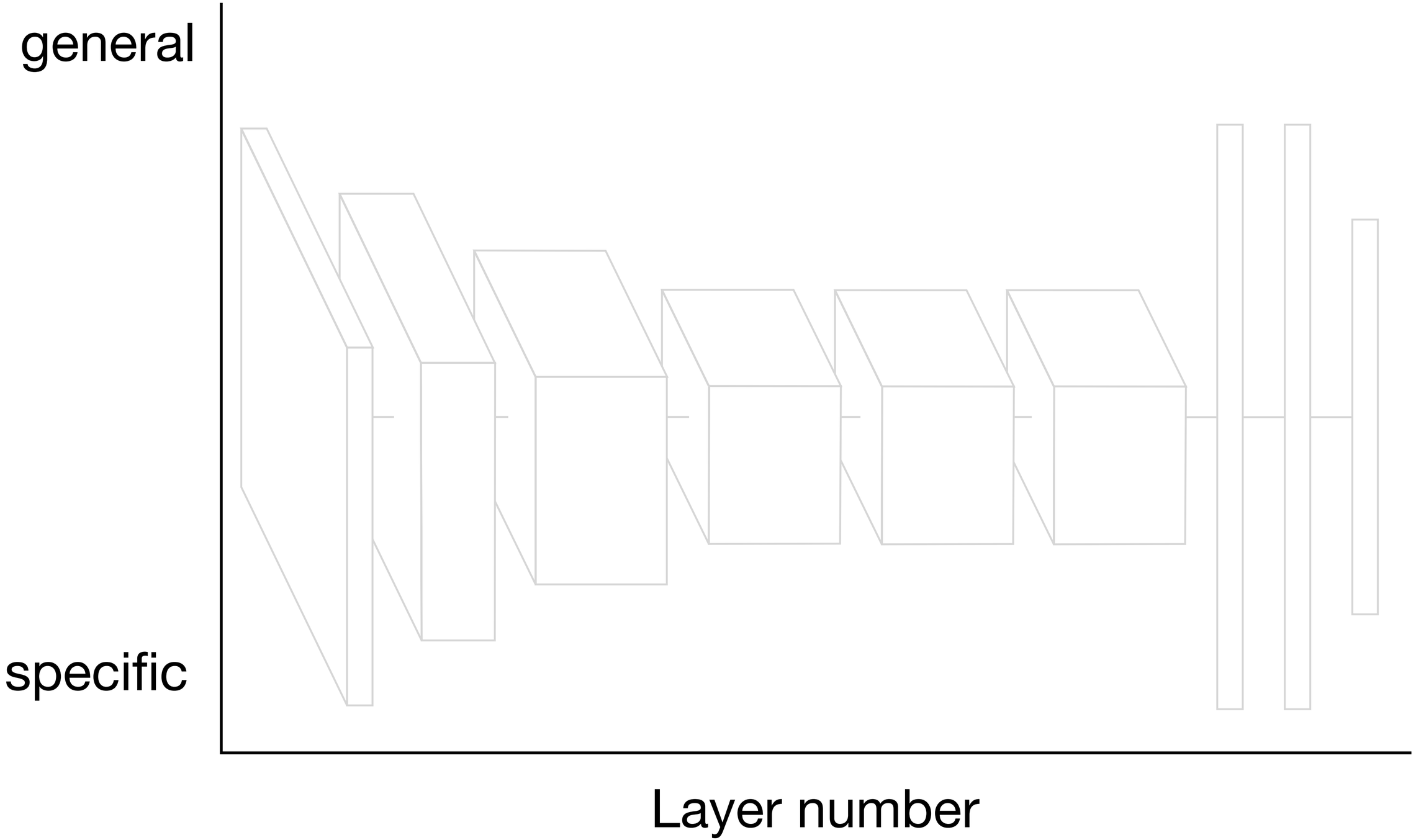
Zeiler et al.
arXiv 2013, ECCV 2014

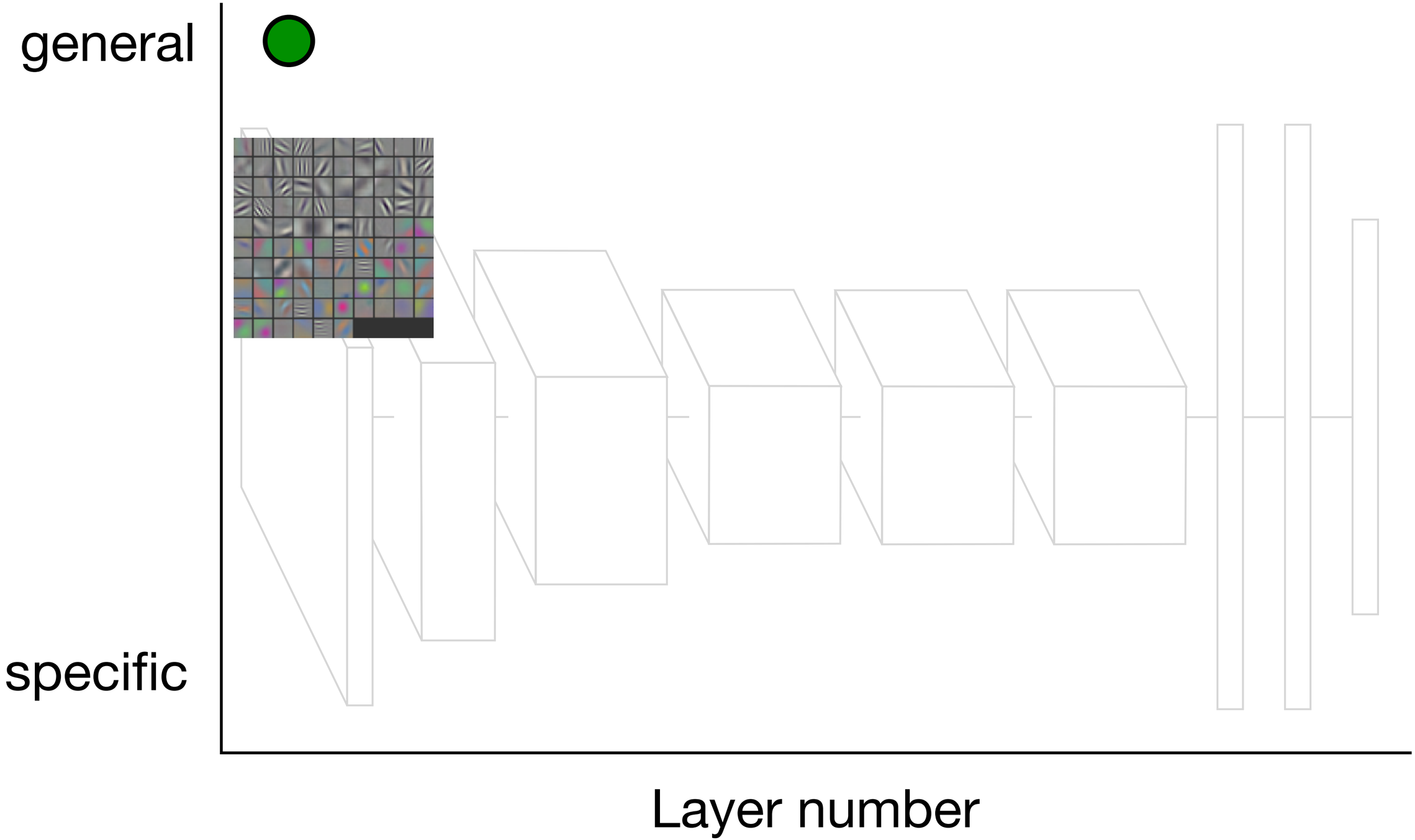


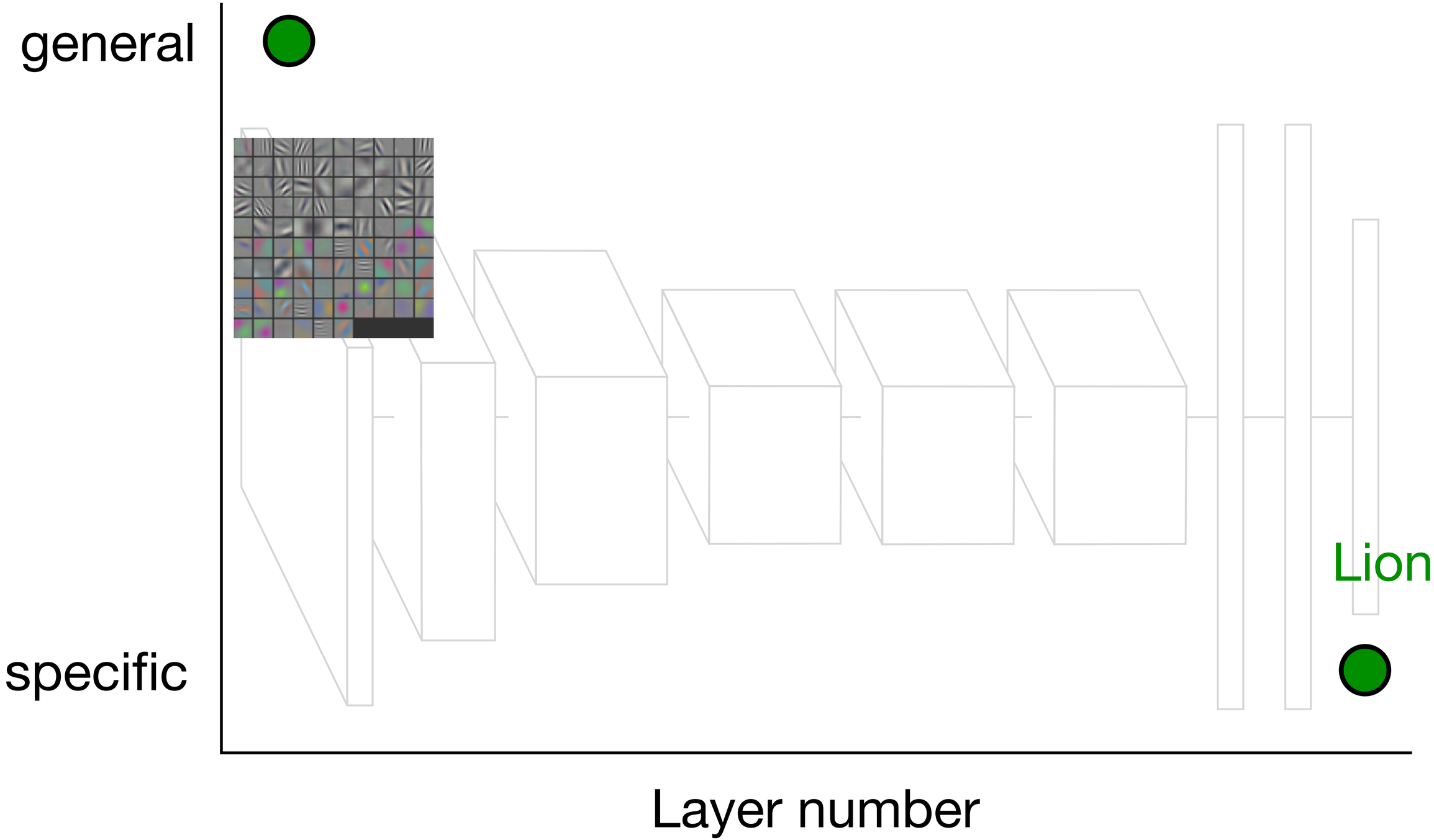
Zeiler et al.
arXiv 2013, ECCV 2014

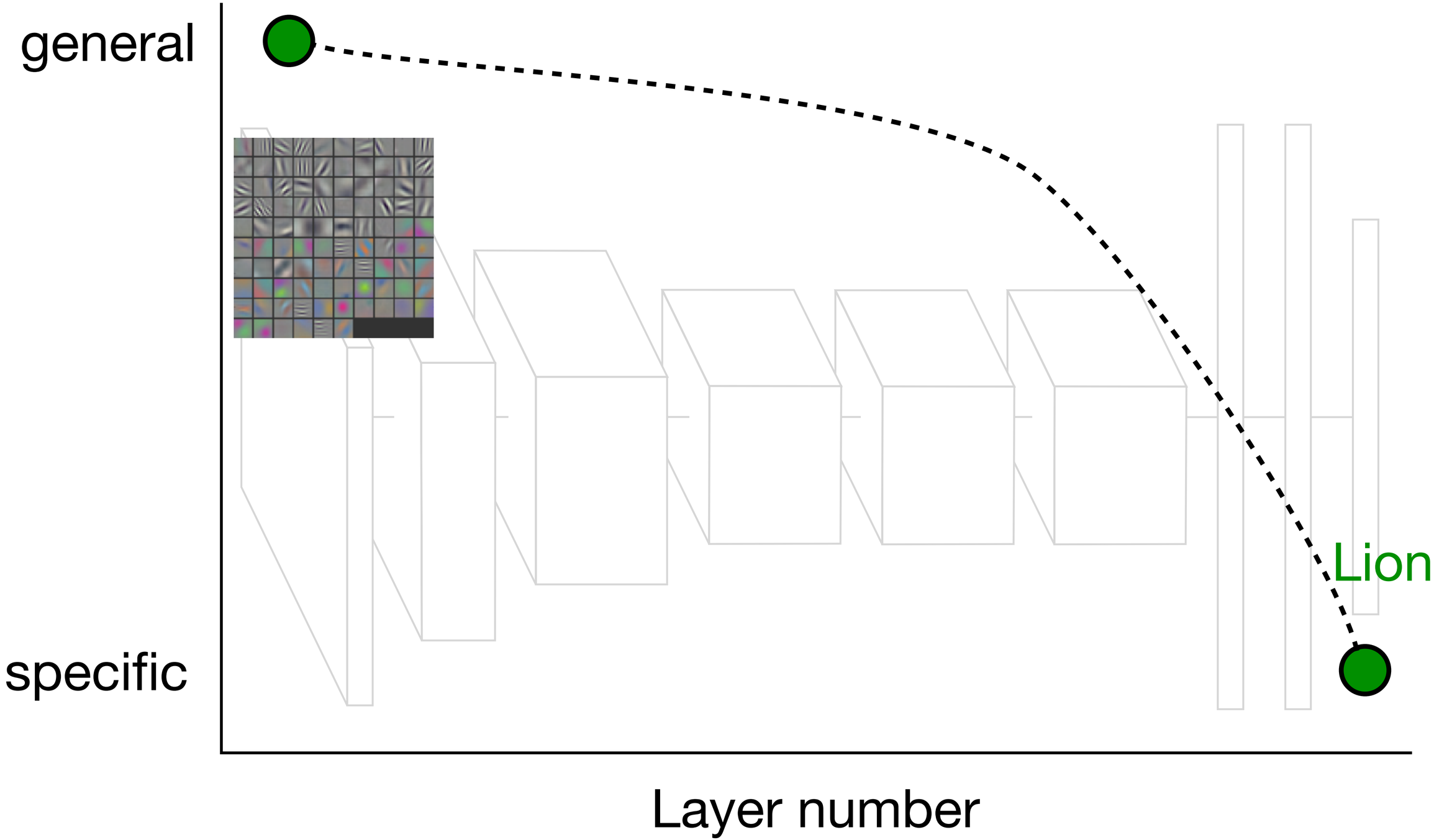
Nguyen et al.
arXiv 2014

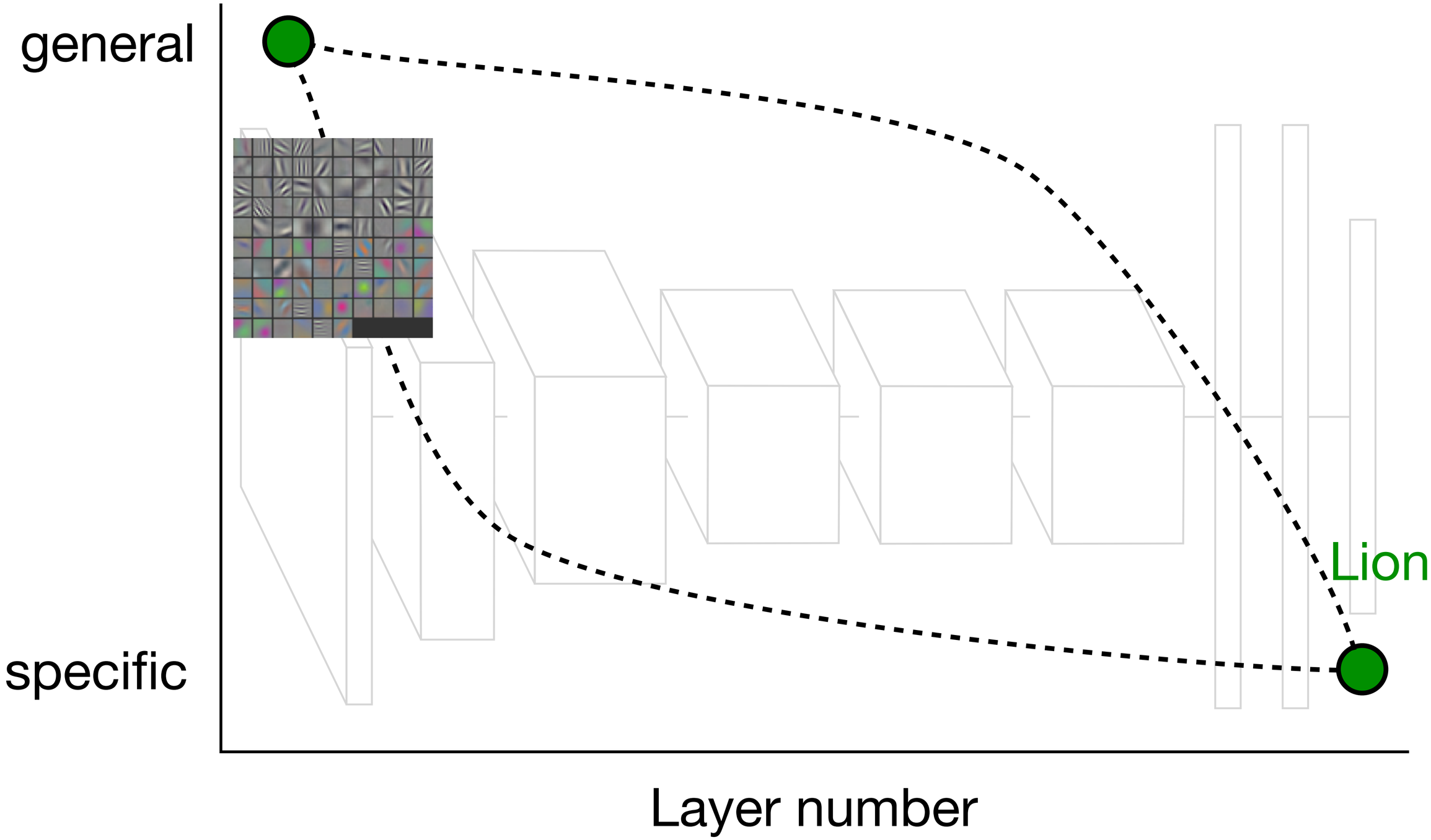


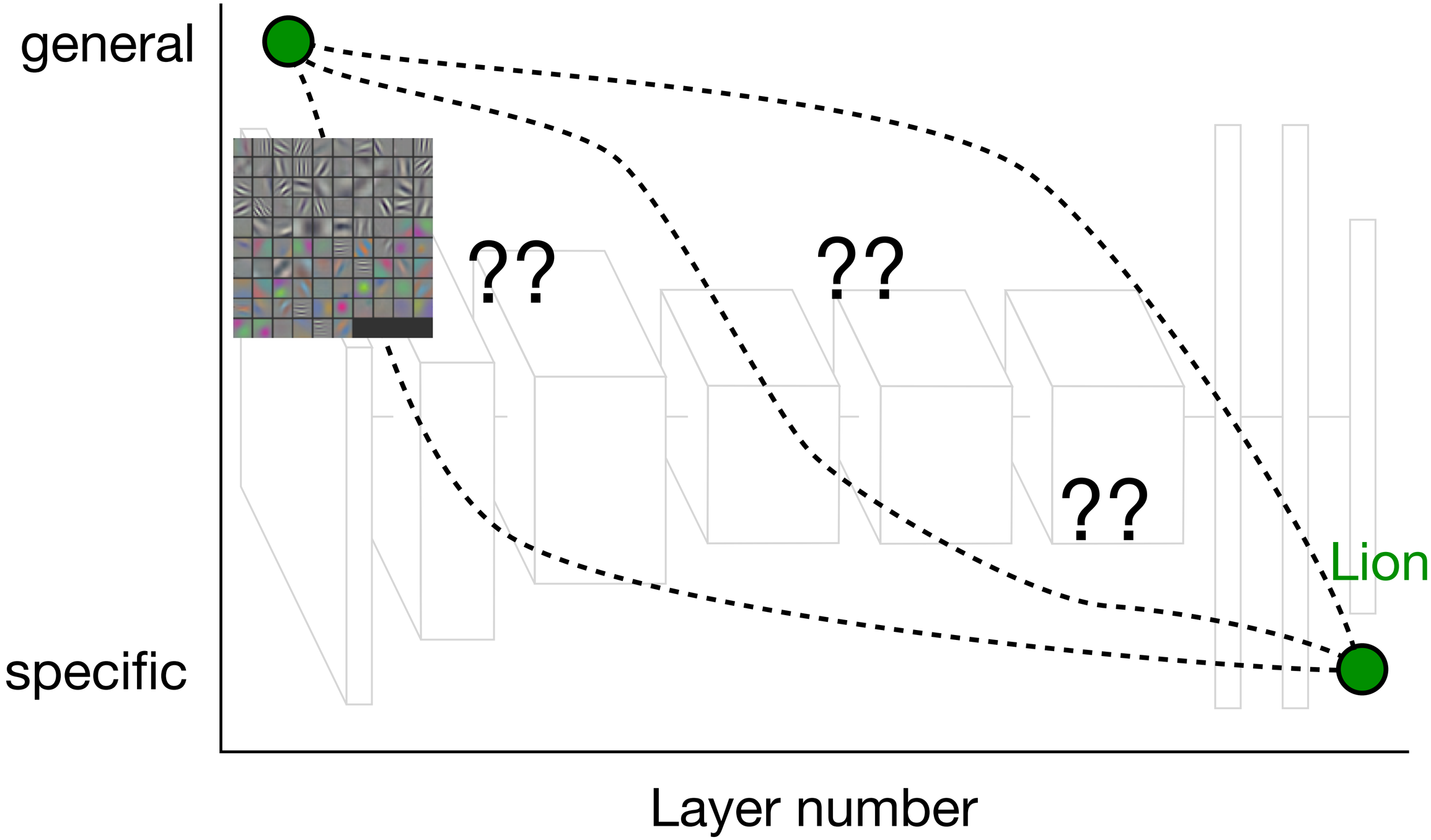

general
specific
Layer number

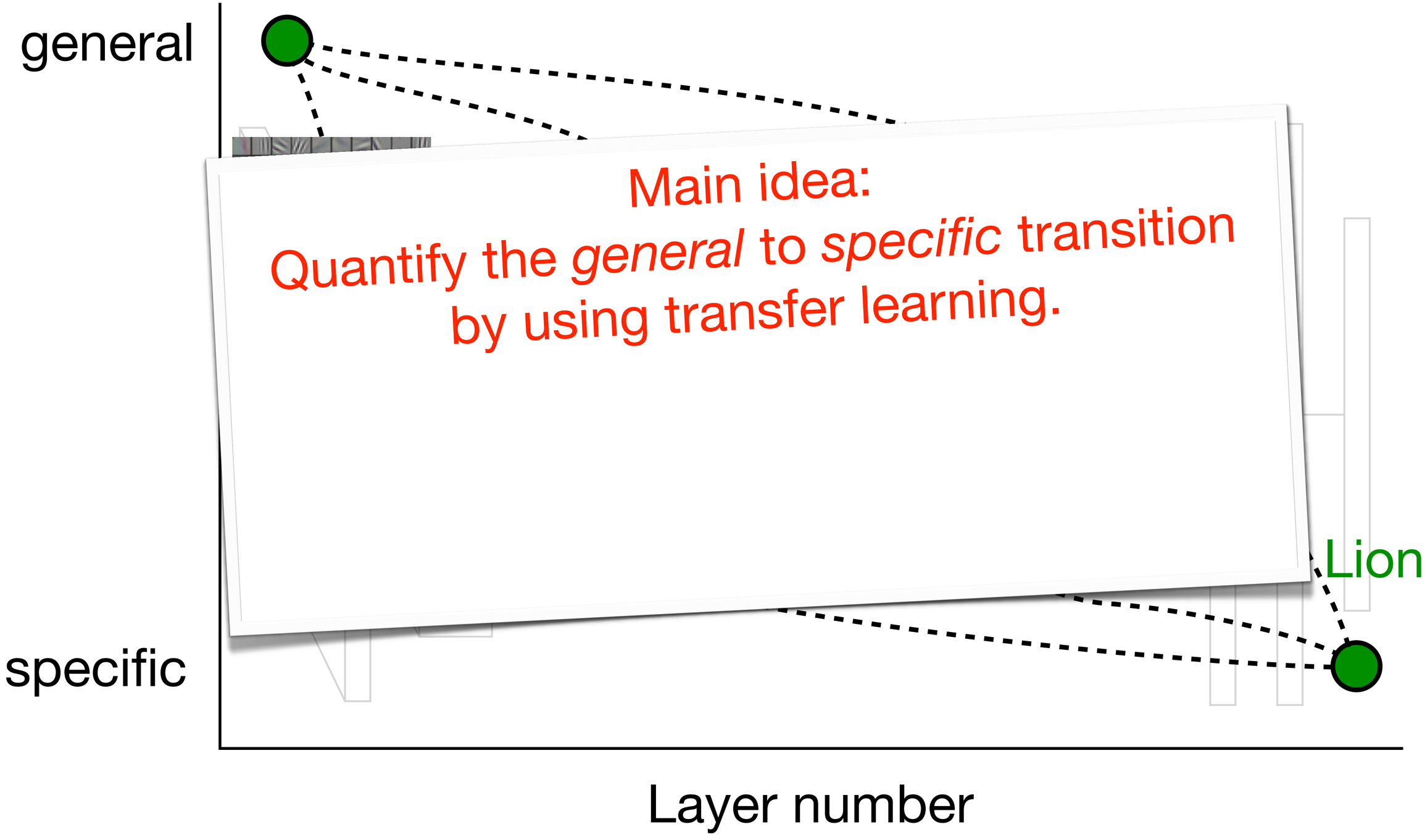

general
specific
Layer number


general
specific
Lion
Layer number

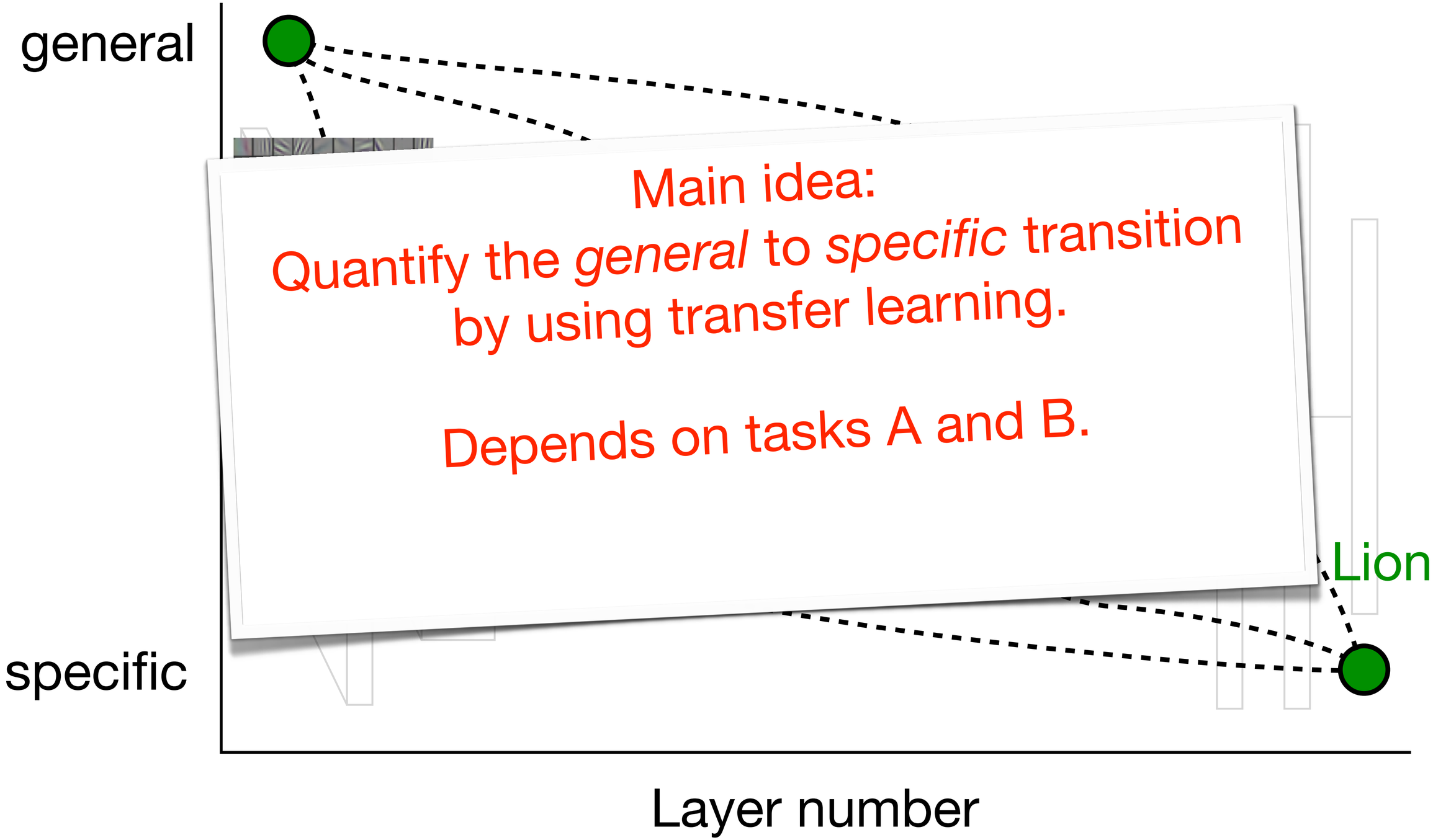

general
specific
Lion
Layer number


general
specific
Lion
Layer number


general
specific
??
??
??
Lion
Layer number

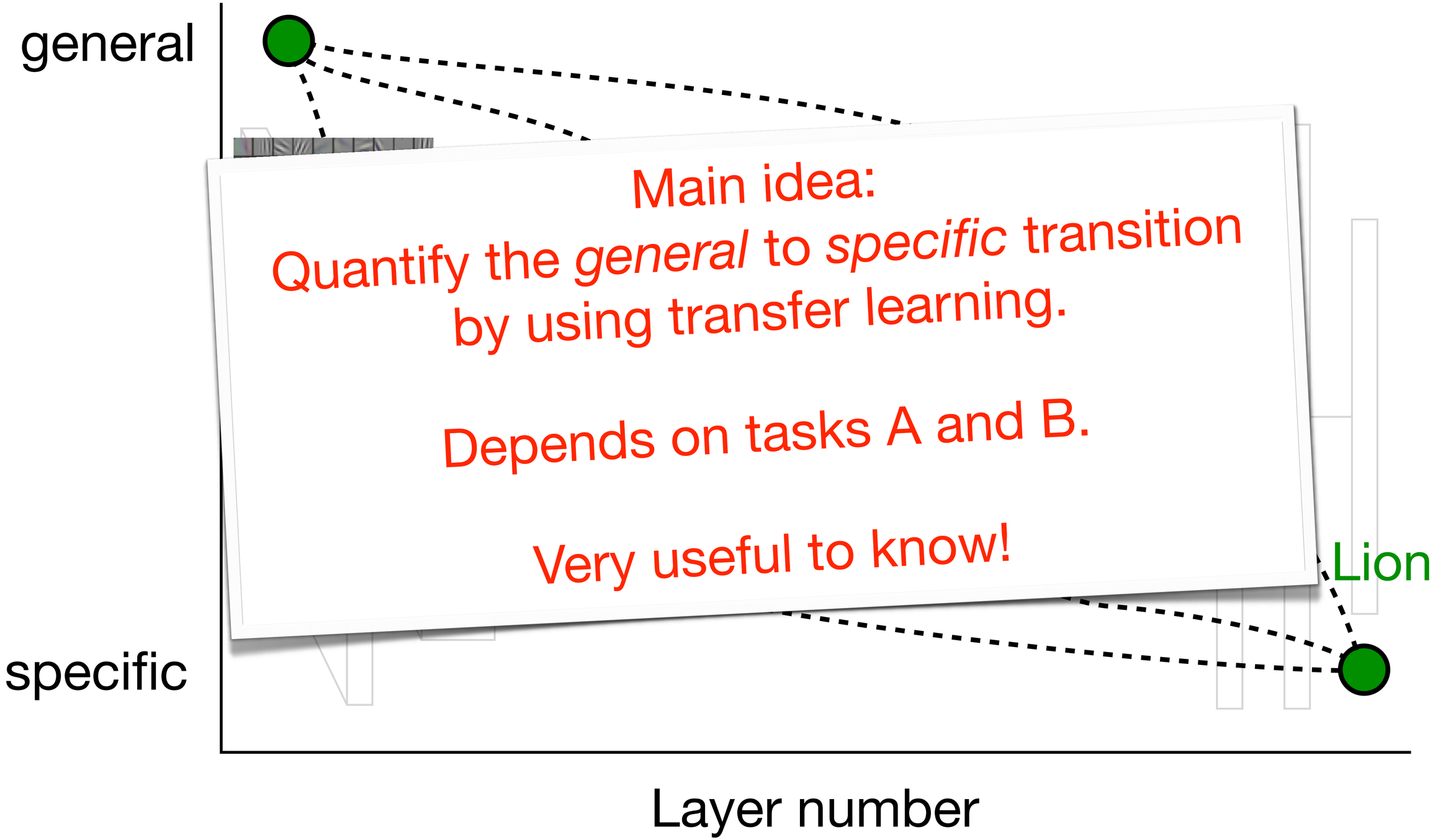


Main idea:
Quantify the *general* to *specific* transition
by using transfer learning.



Main idea:
Quantify the *general* to *specific* transition by using transfer learning.

Depends on tasks A and B.



Main idea:
Quantify the *general* to *specific* transition by using transfer learning.

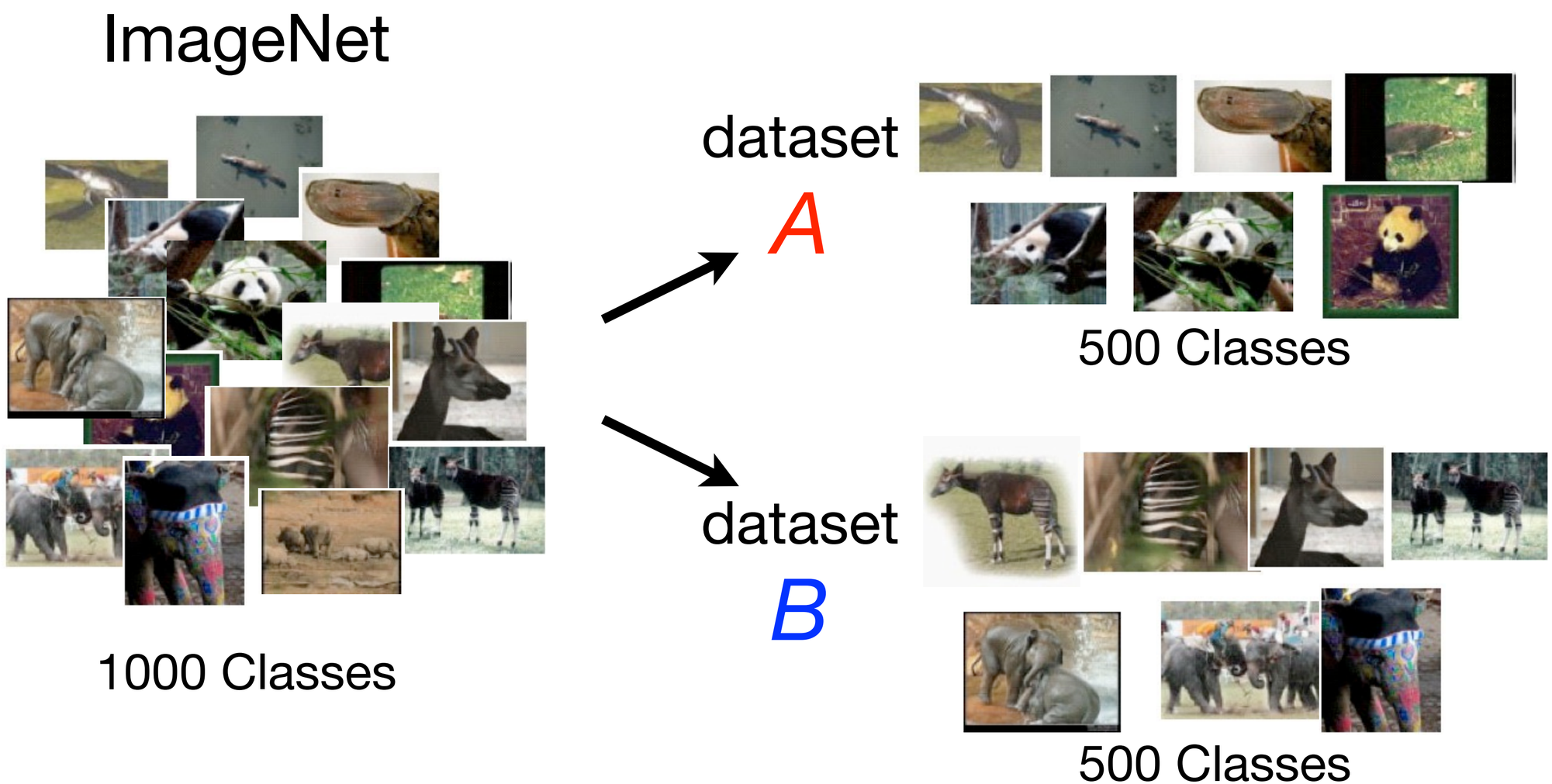
Depends on tasks A and B.

Very useful to know!

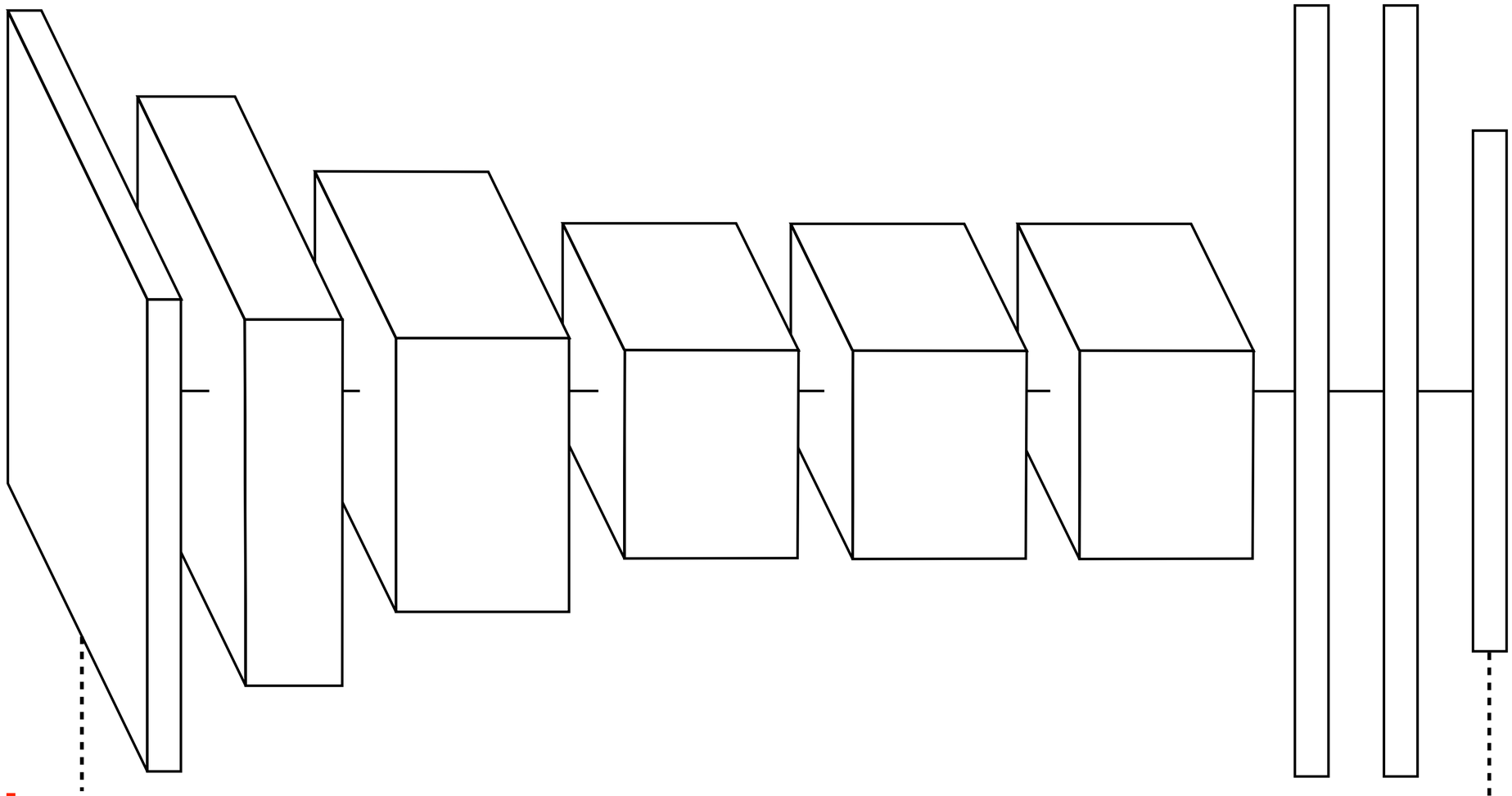
ImageNet

1000 Classes

Deng et al., 2009

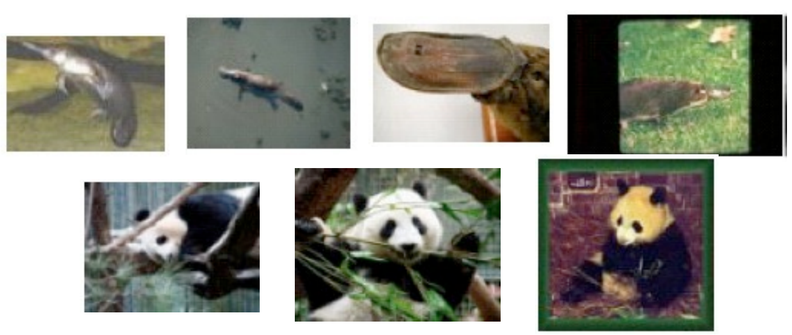


Deng et al., 2009



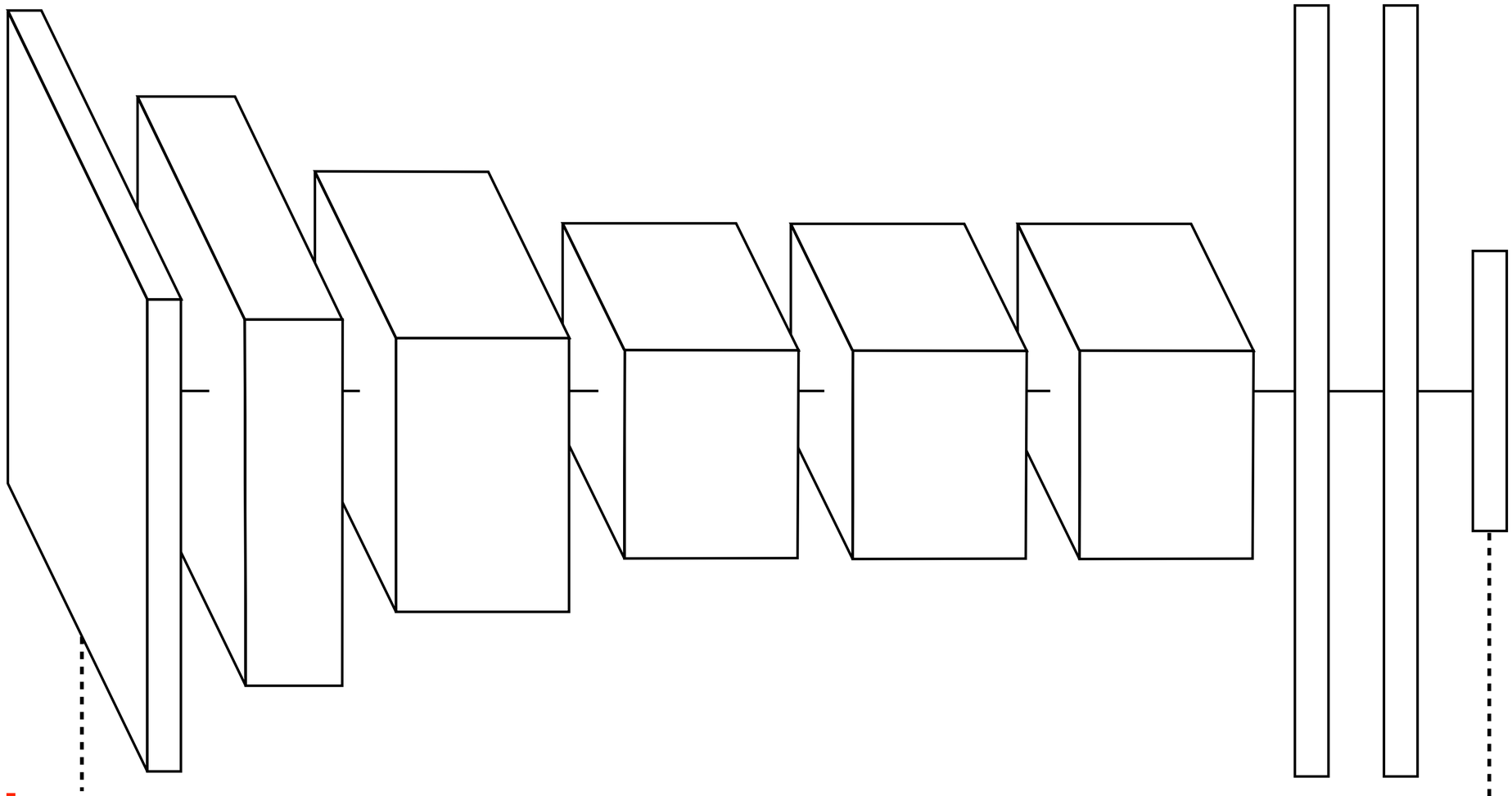
A Images

A Labels



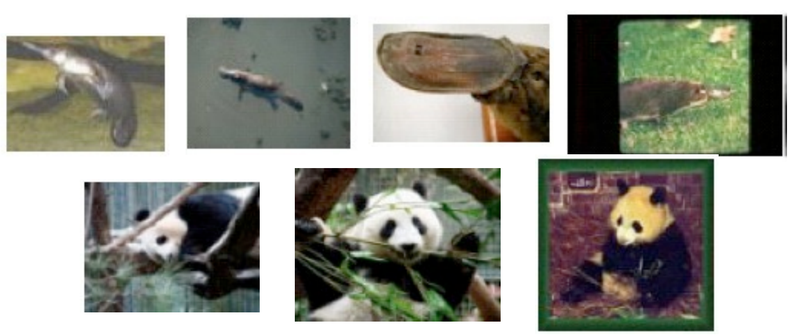
500 Classes

Train using Caffe framework (Jia et al.)



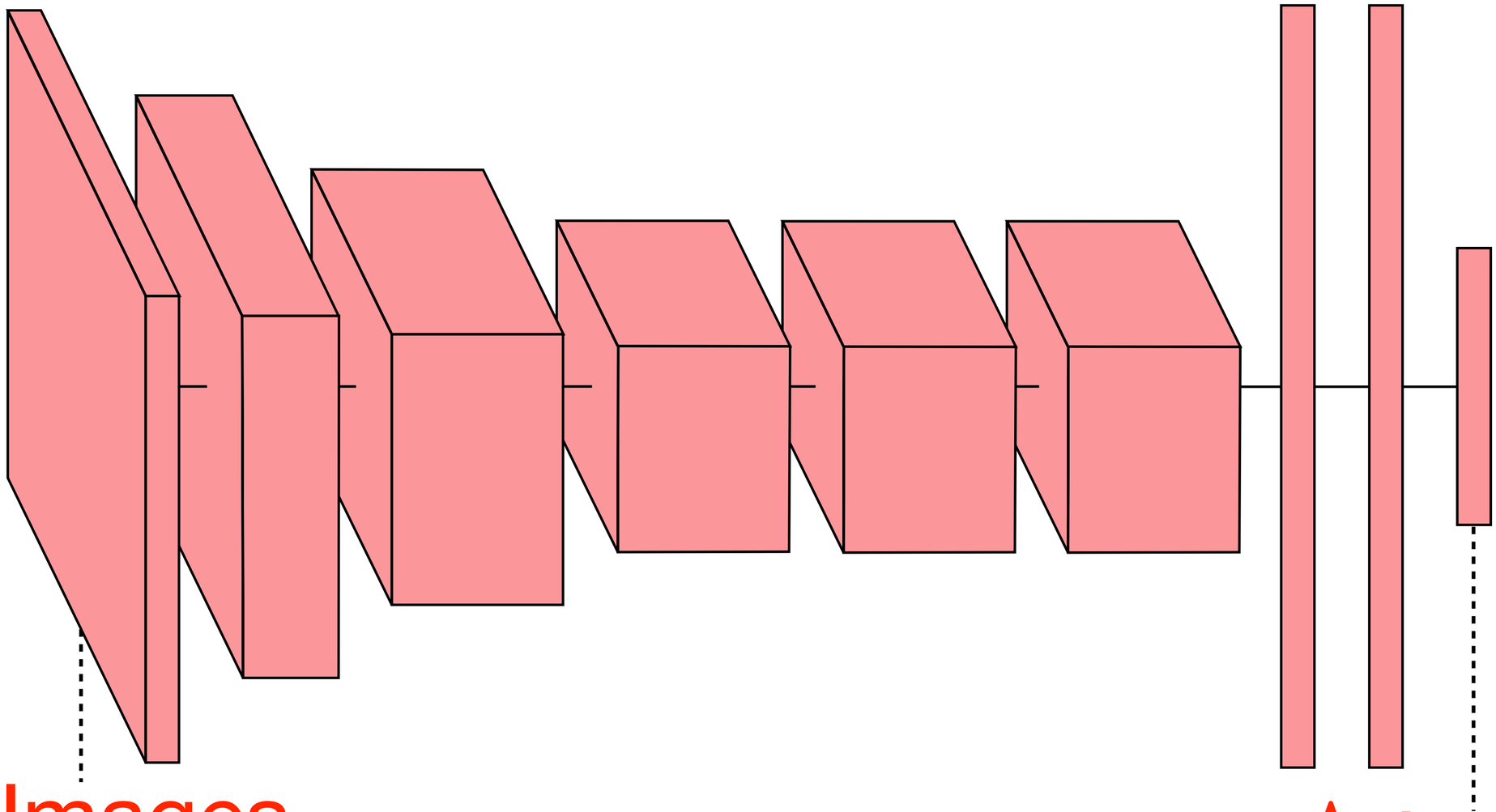
A Images

A Labels



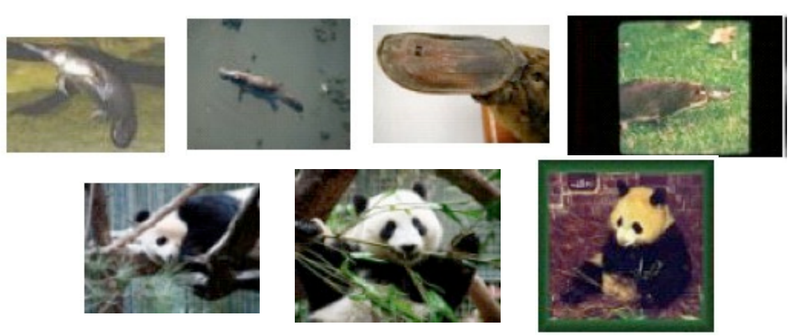
500 Classes

Train using Caffe framework (Jia et al.)



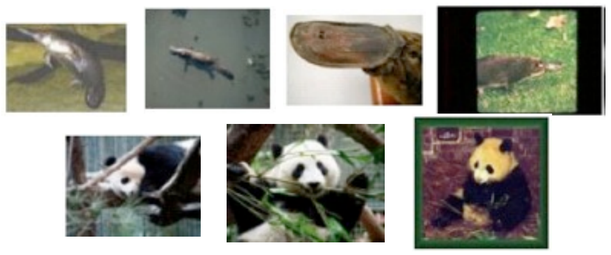
A Images

A Labels

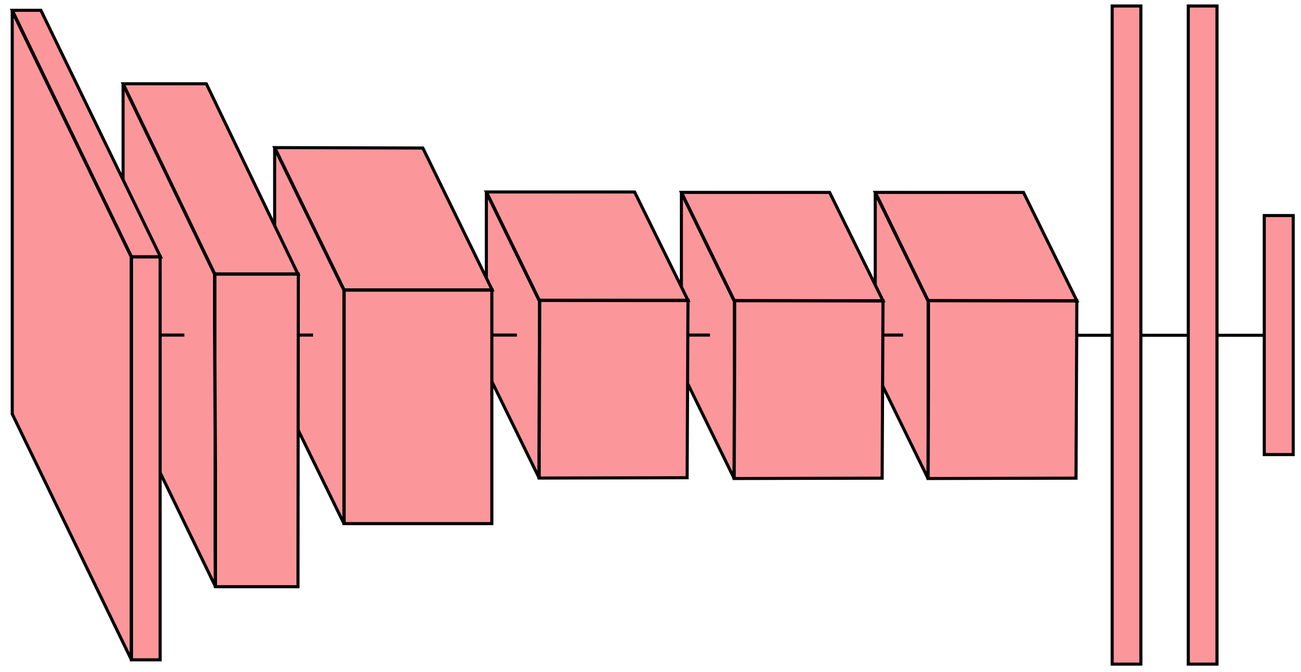


500 Classes

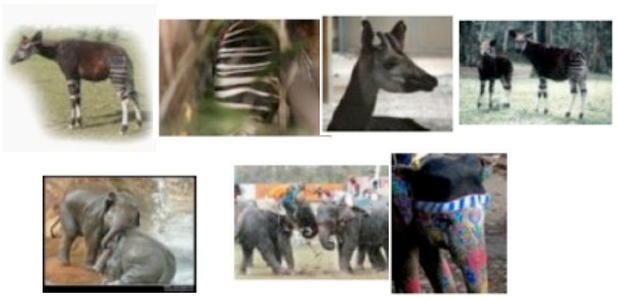
Train using Caffe framework (Jia et al.)



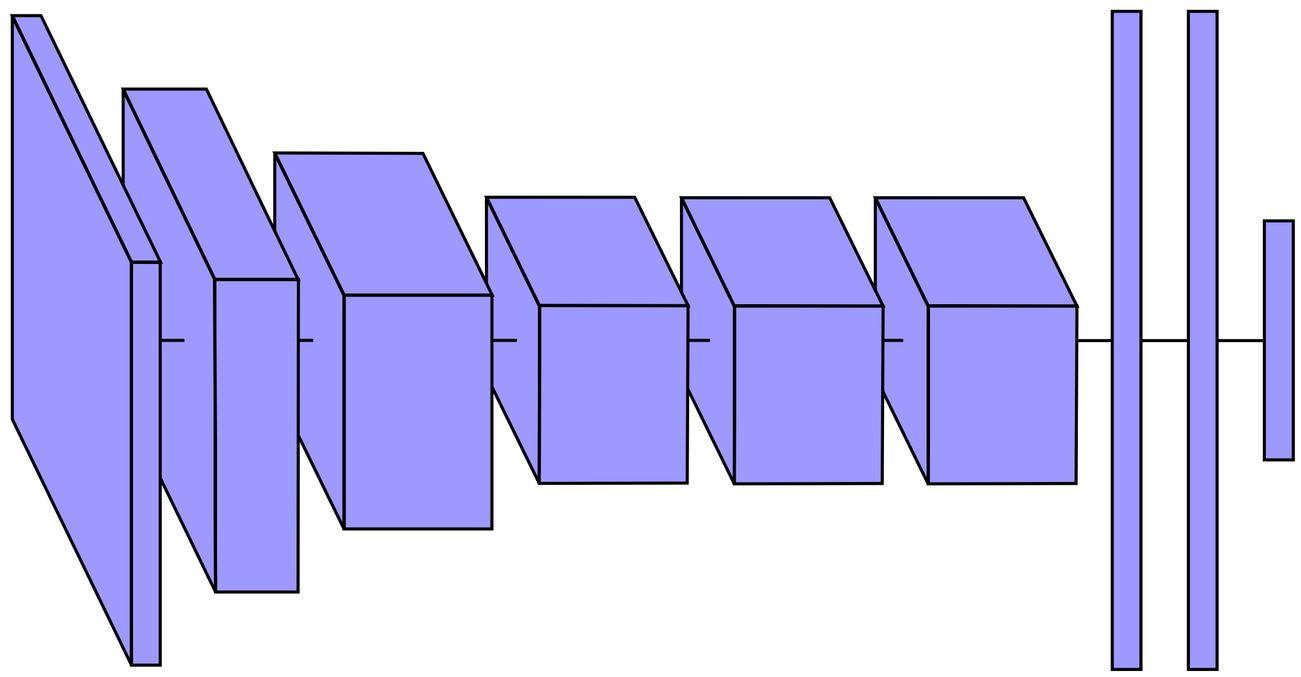
A Images



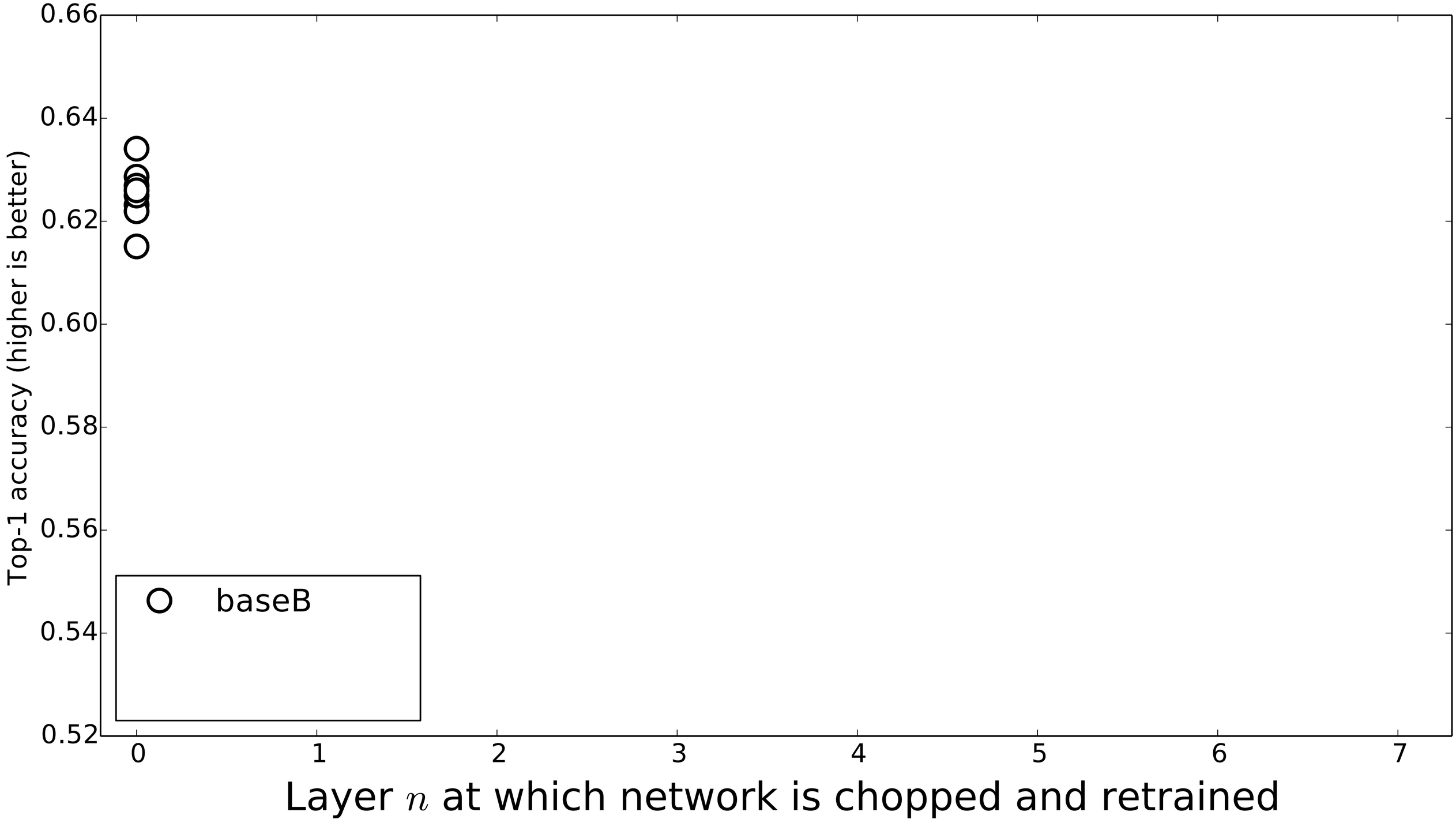
baseA

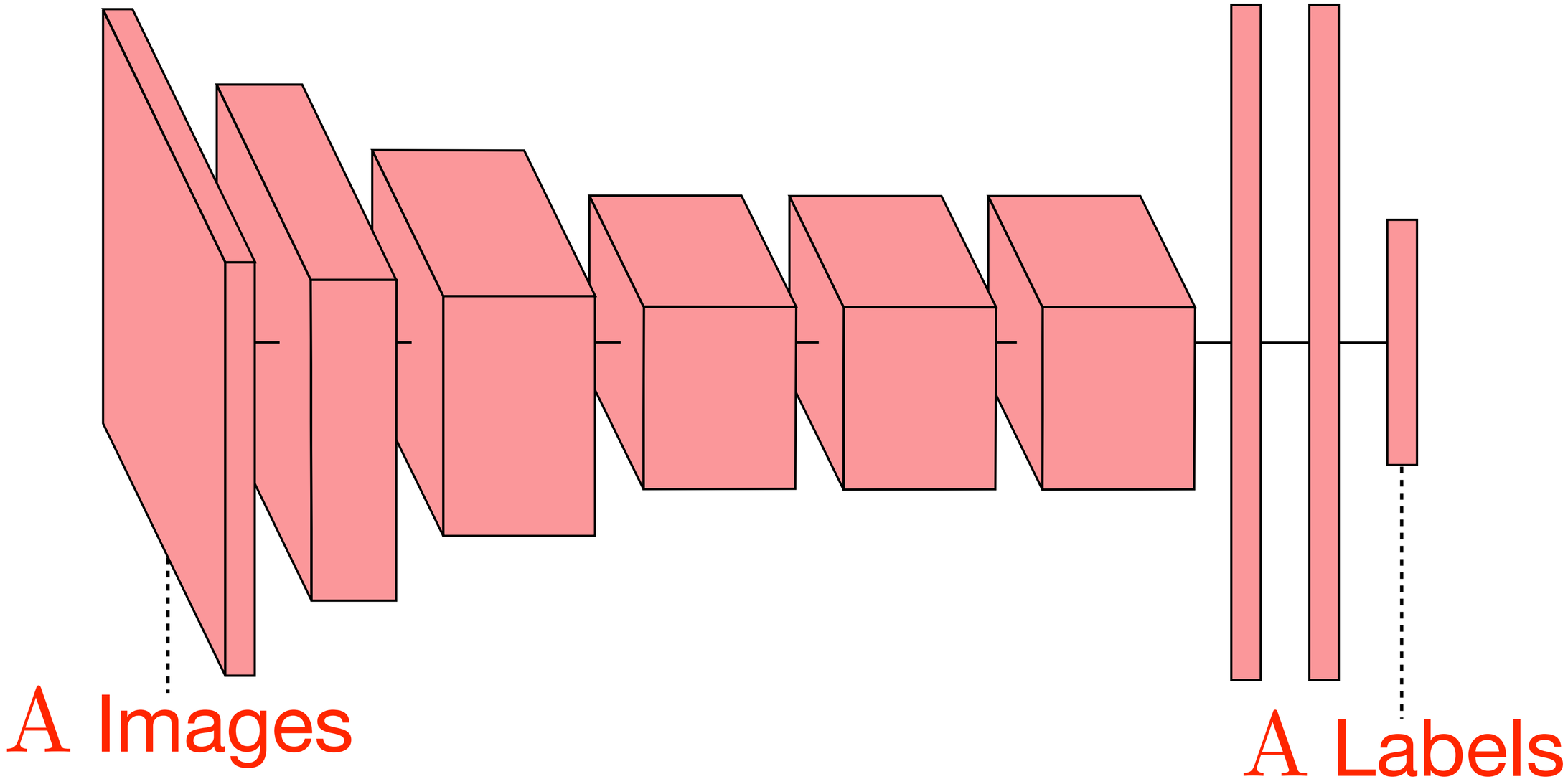


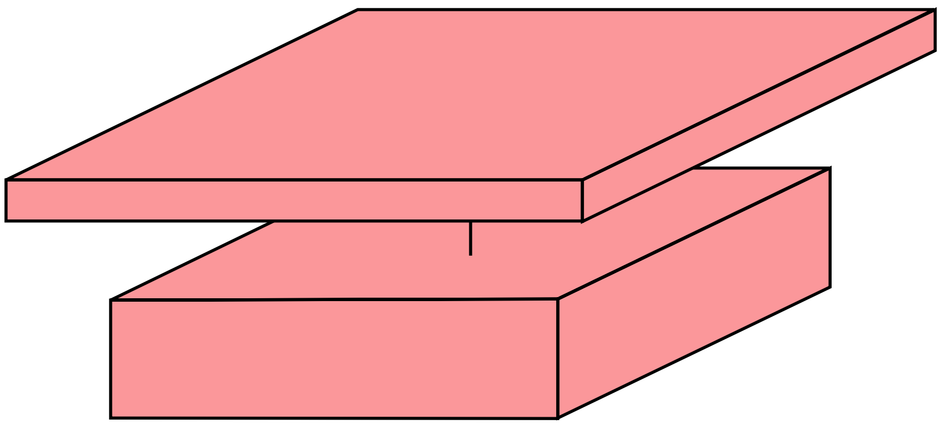
B Images

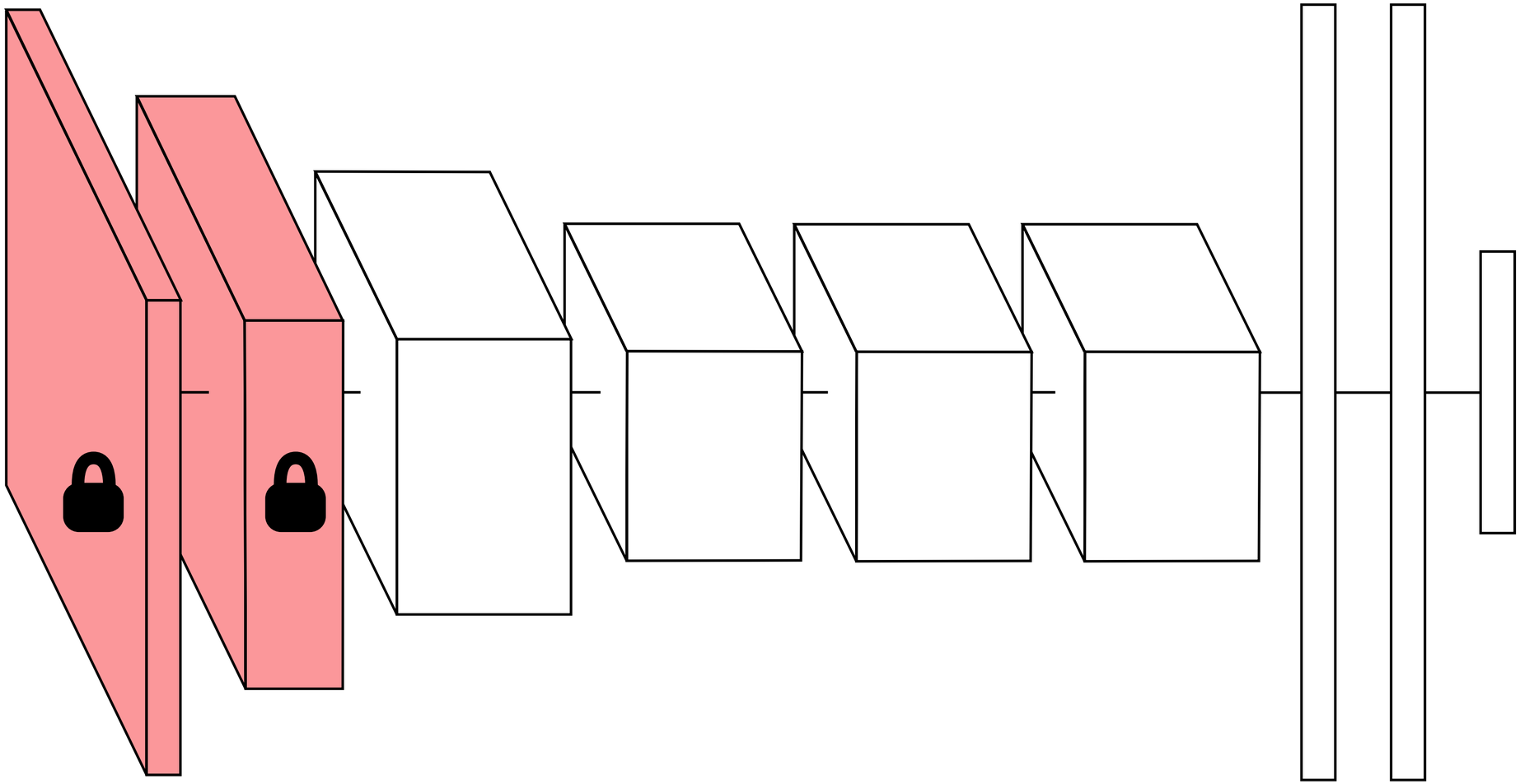


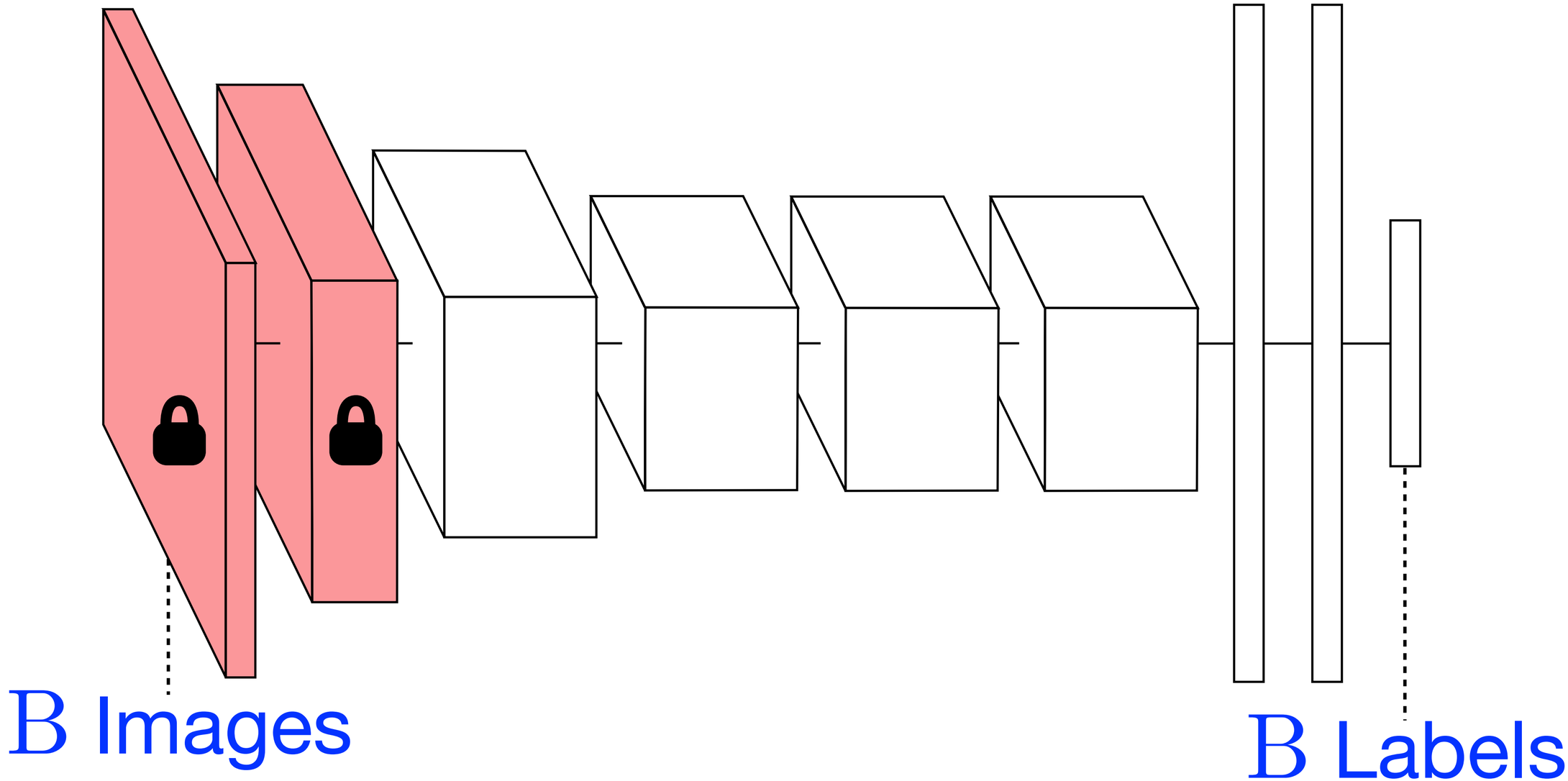
baseB

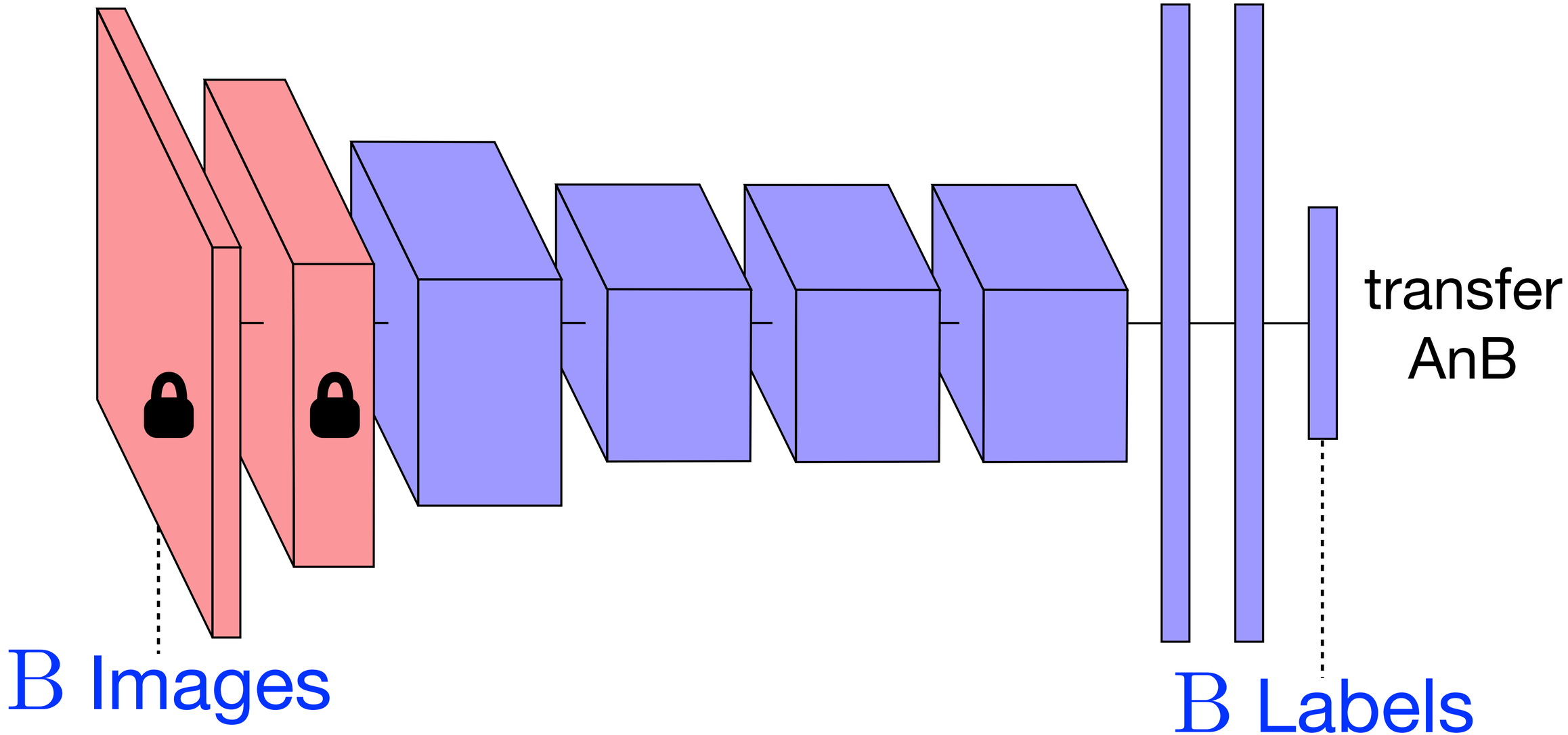

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
baseB
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained

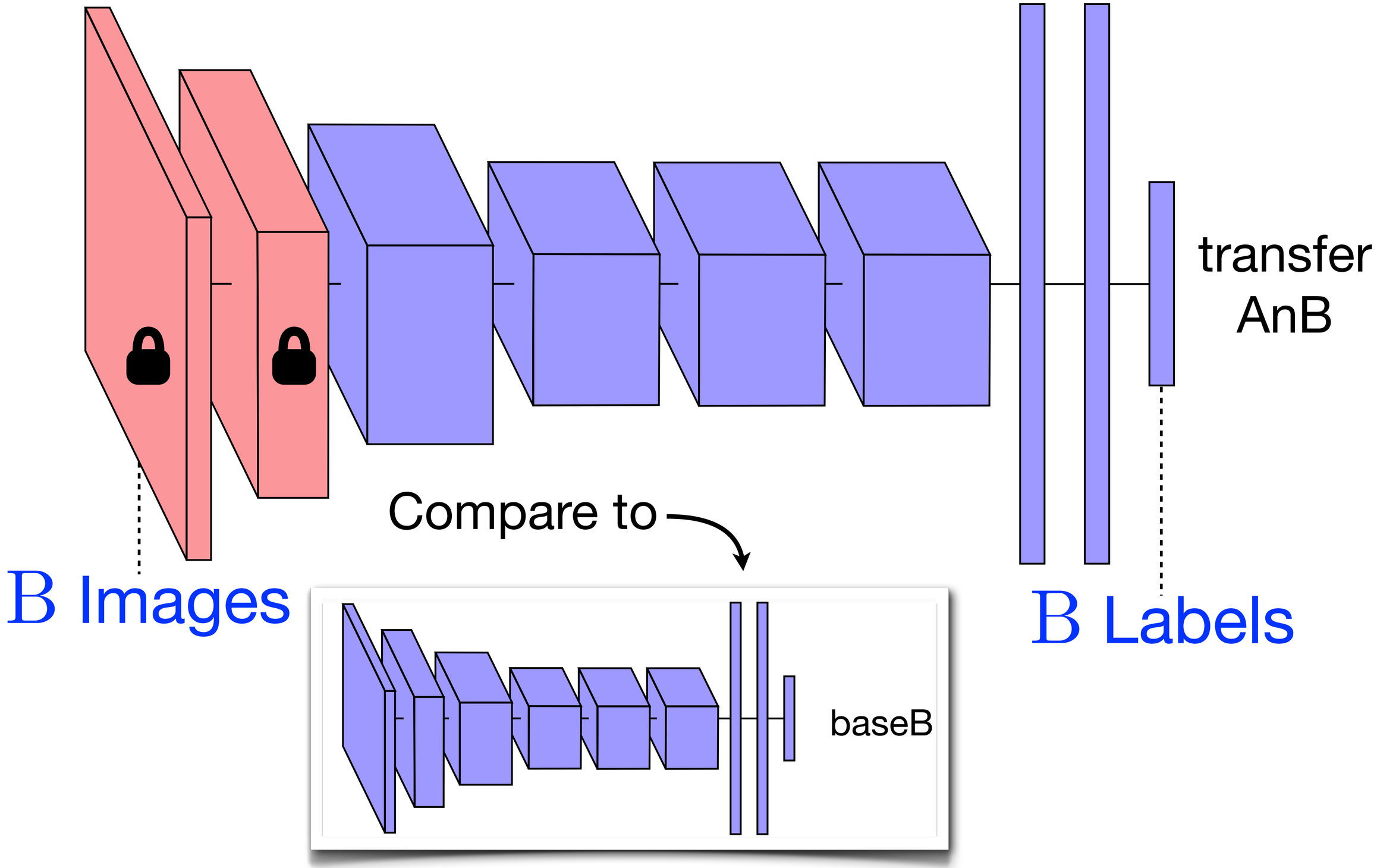

A Images
A Labels

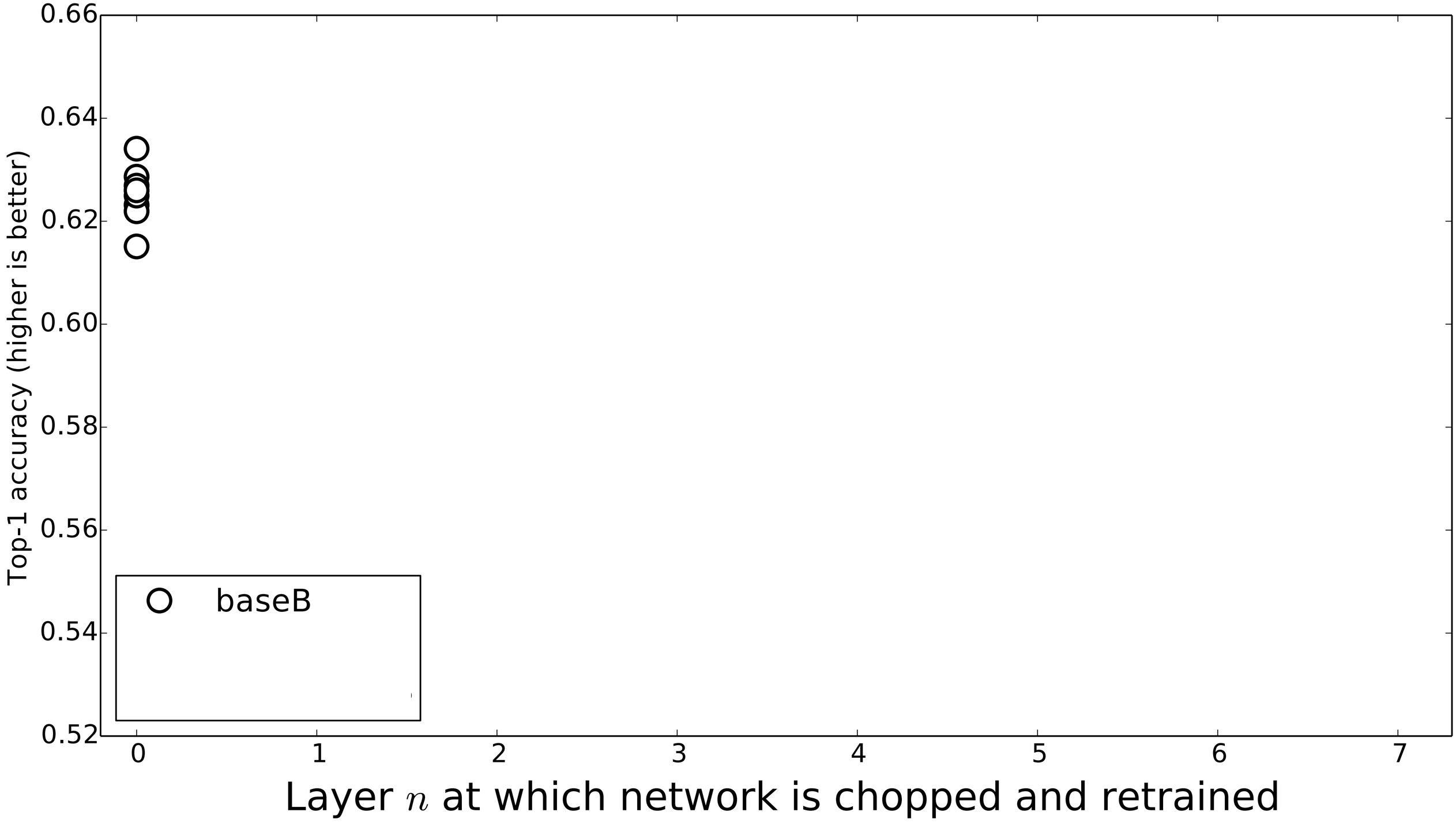


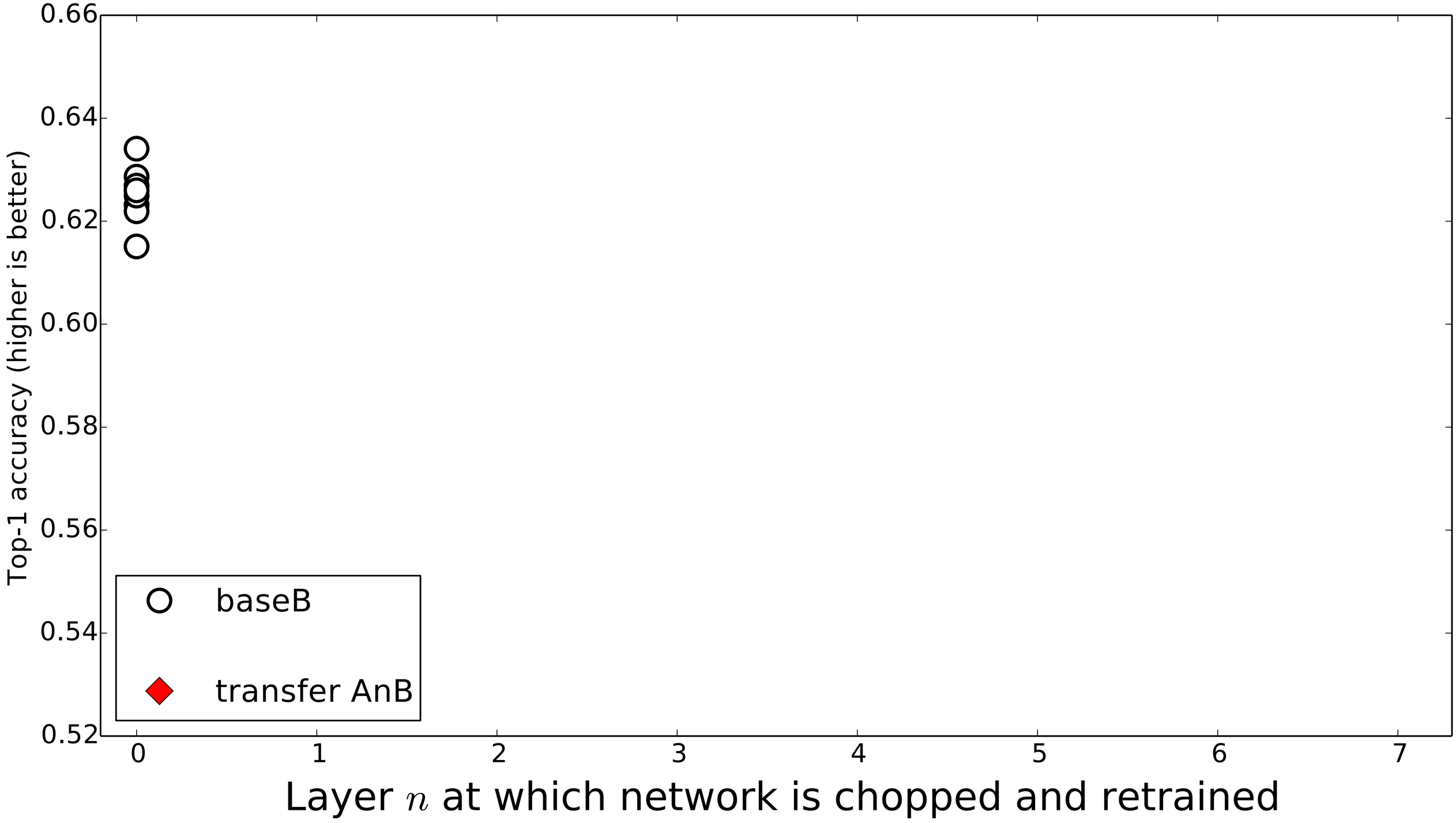


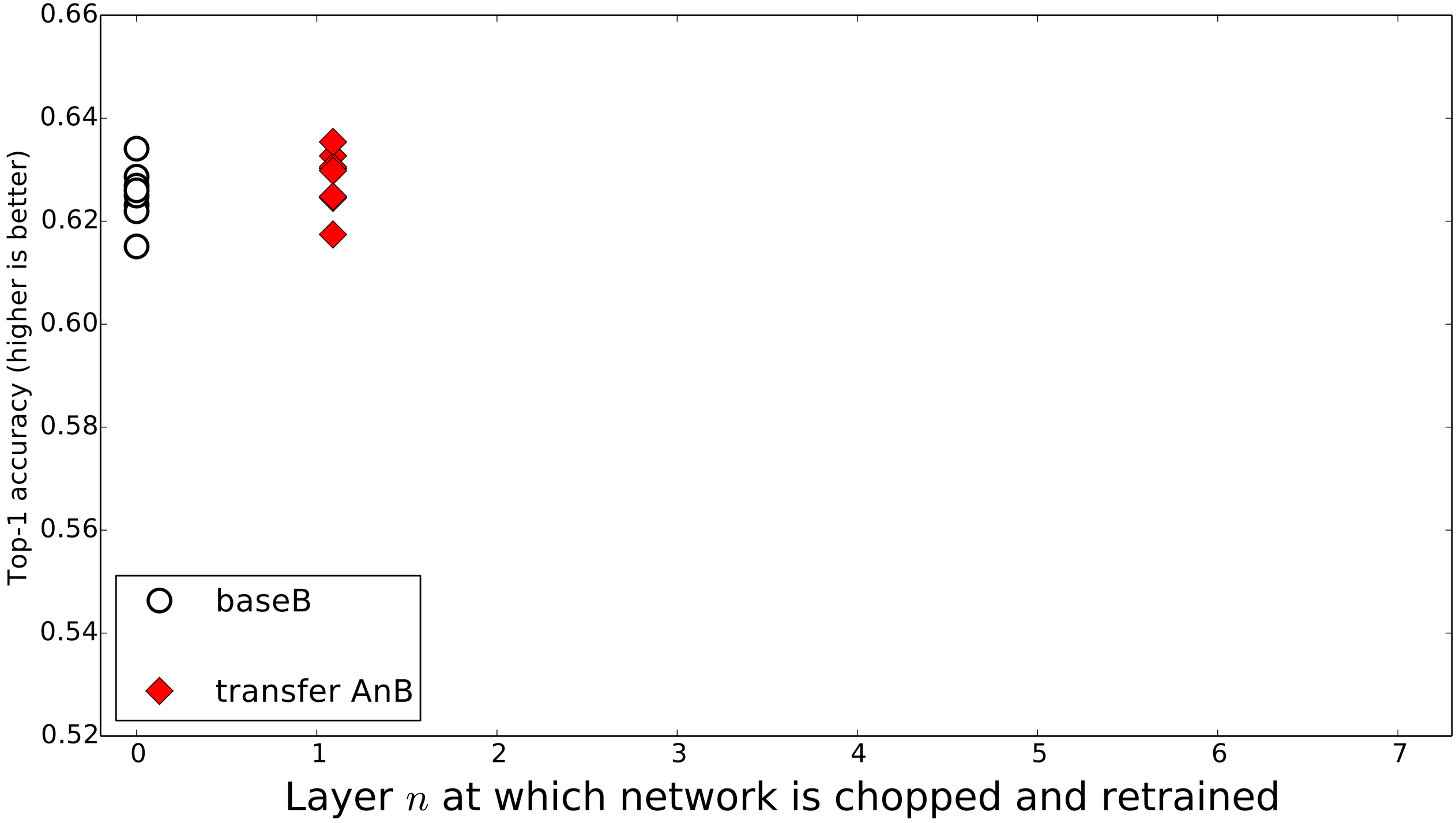

B Images
B Labels

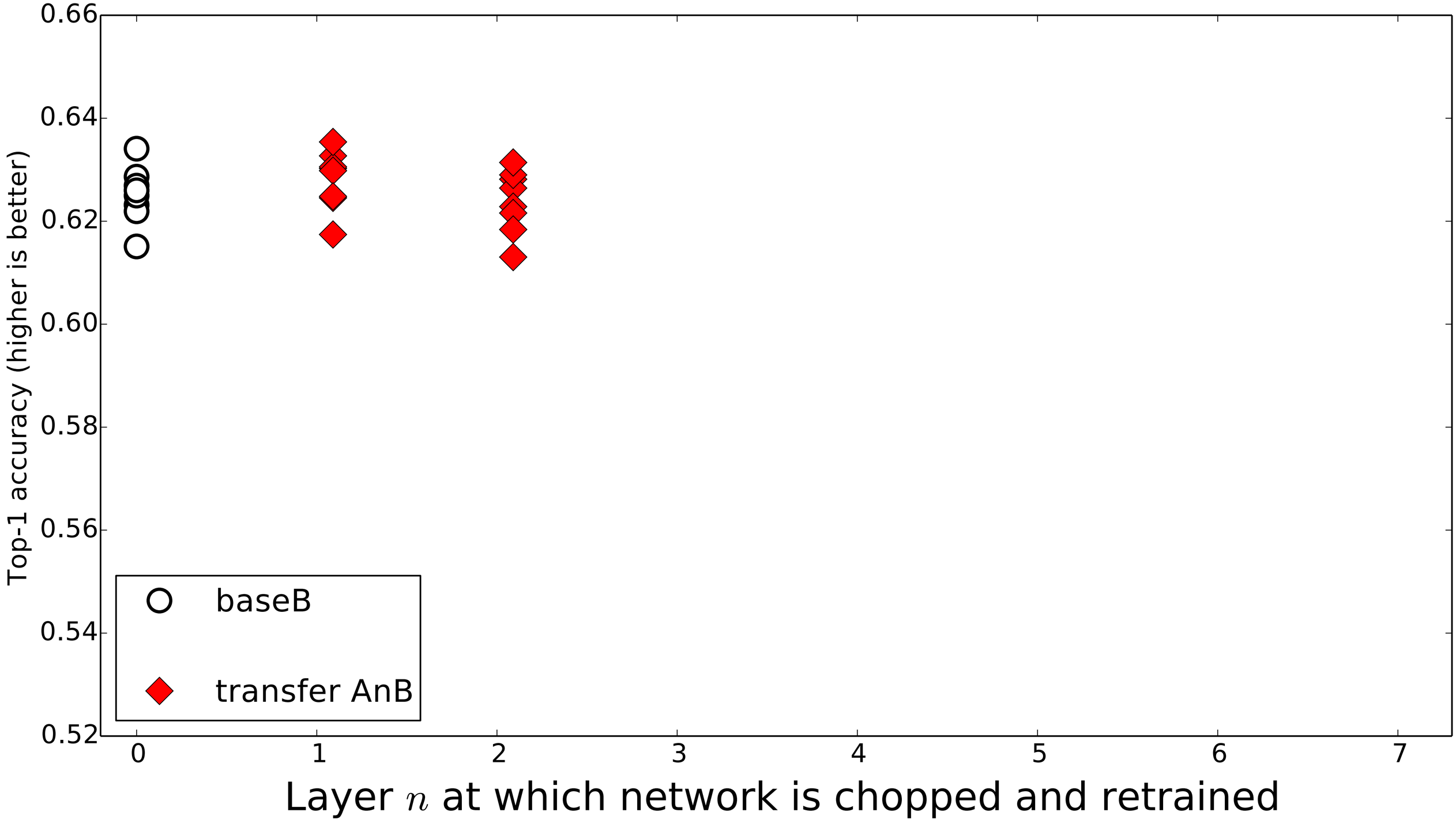

transfer
AnB
B Images
B Labels

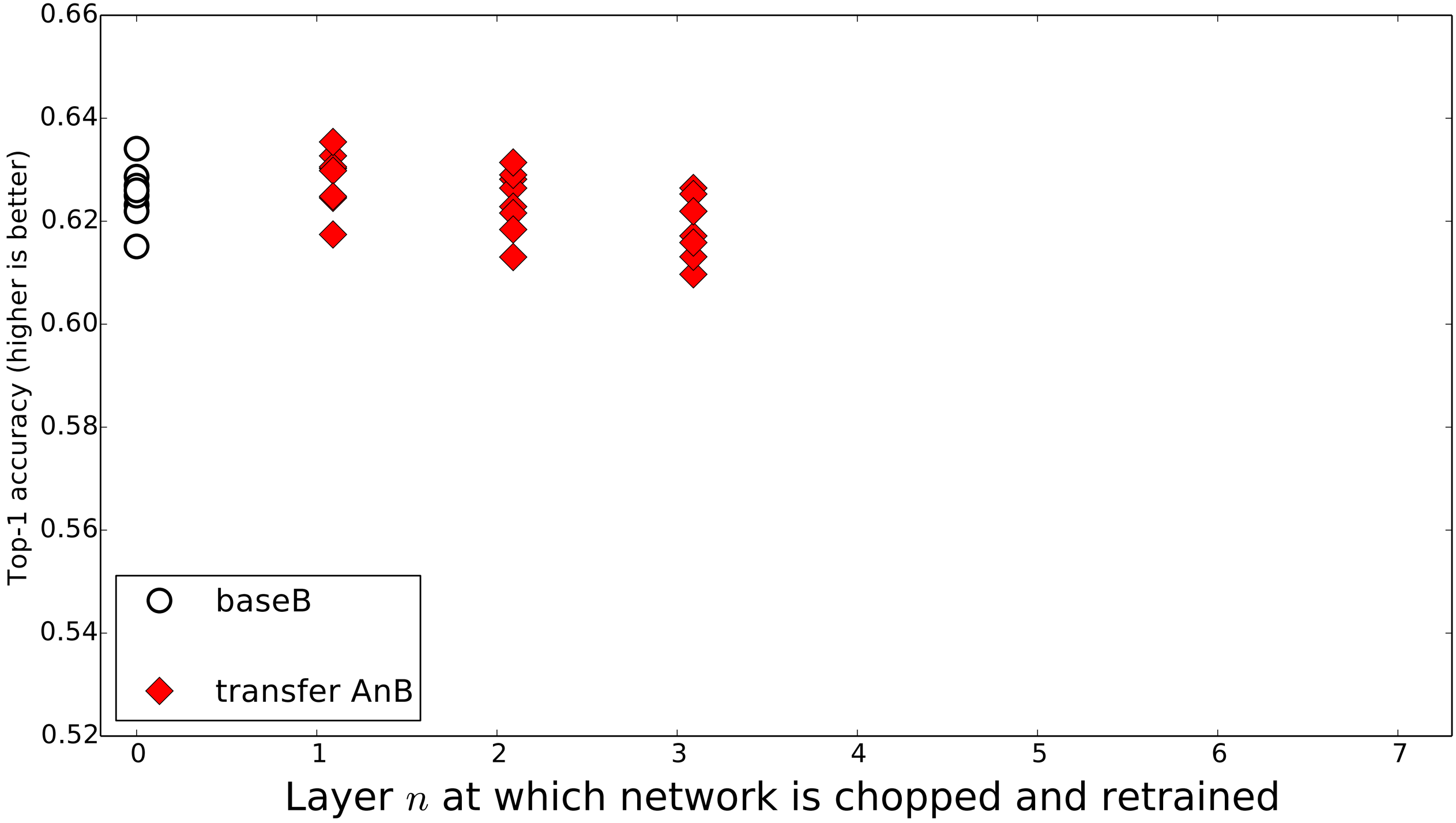

transfer
AnB
Compare to
B Images
B Labels
baseB

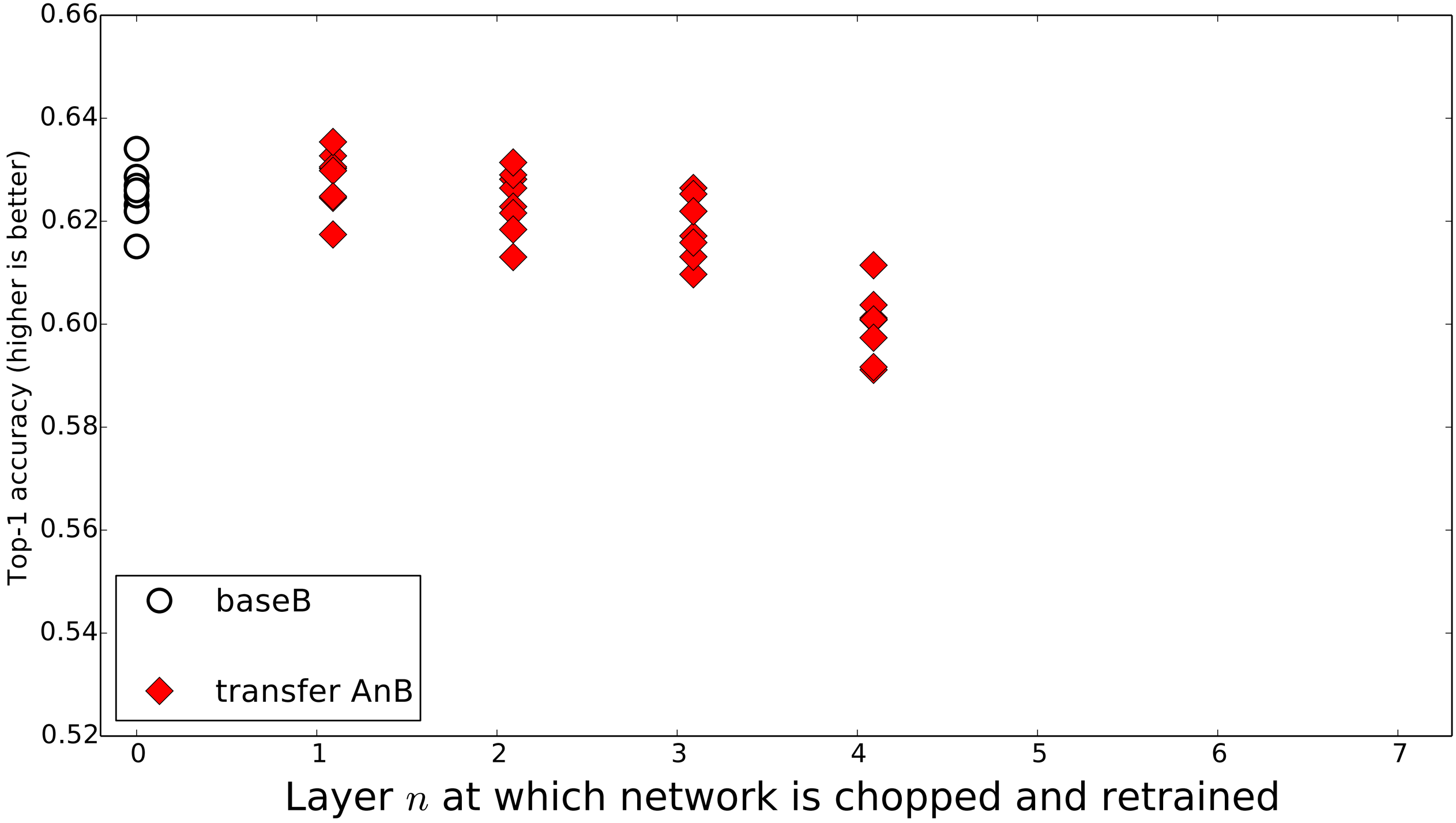

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
baseB
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained

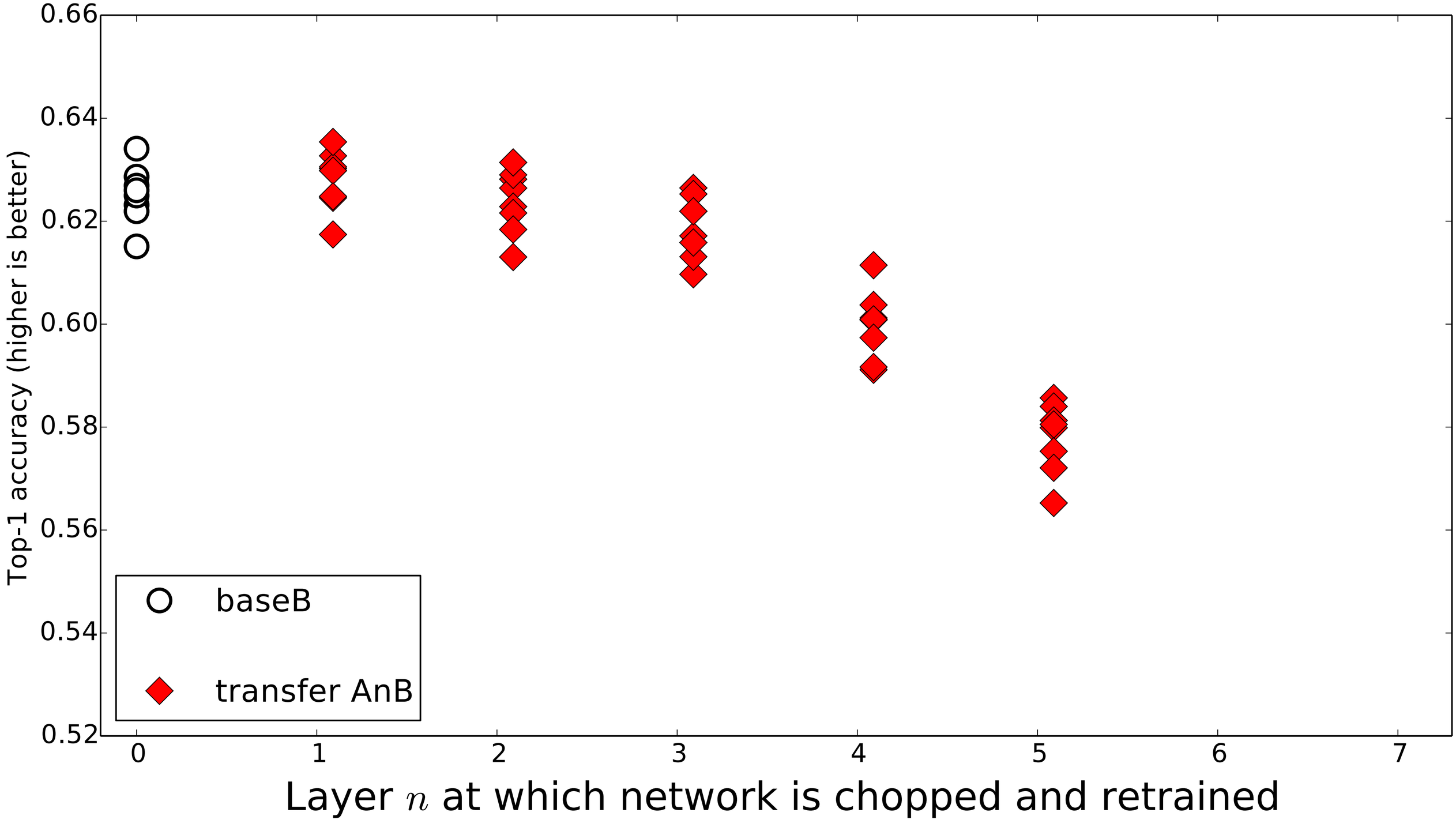

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
transfer AnB

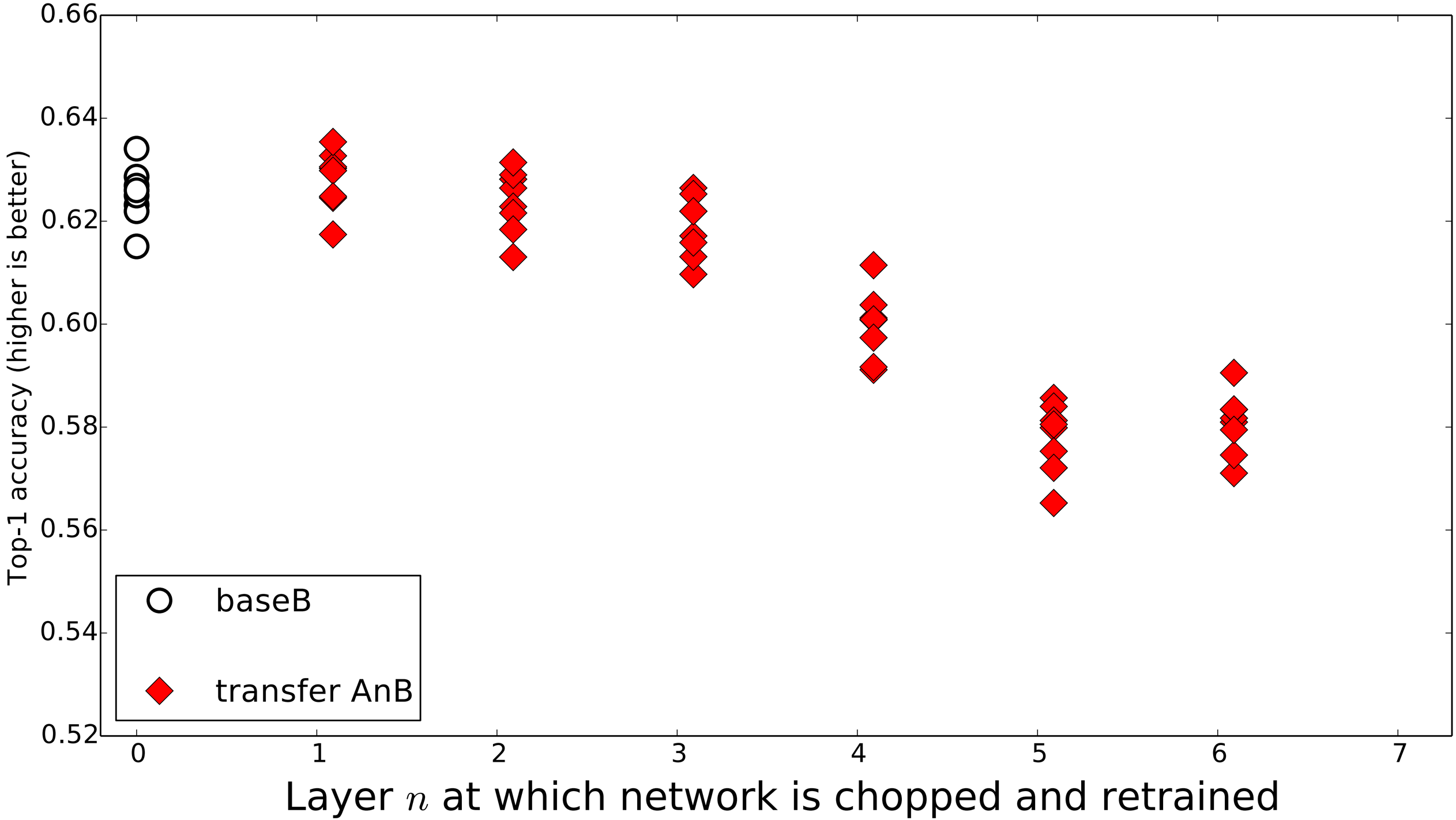

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
transfer AnB

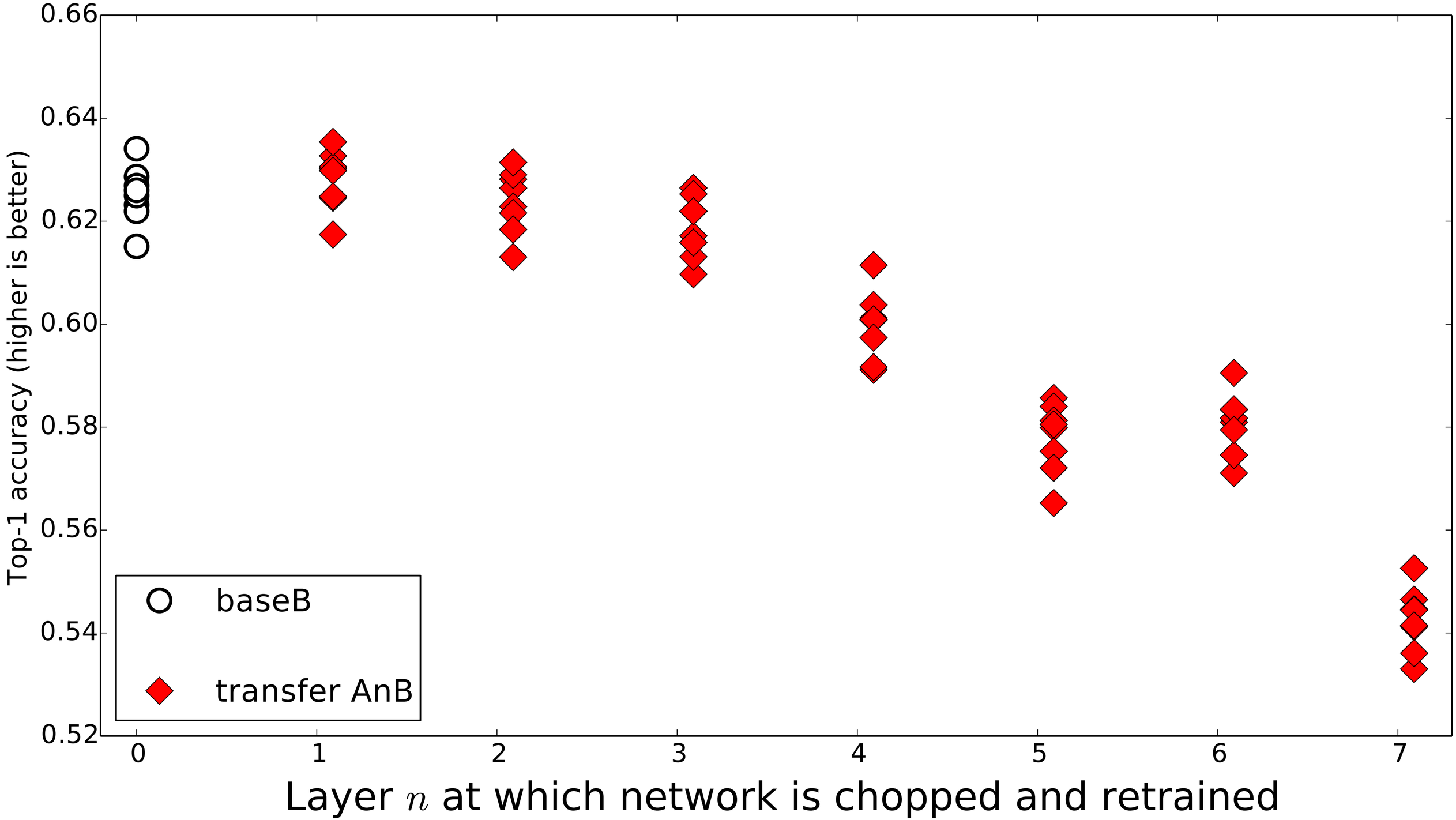

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
transfer AnB

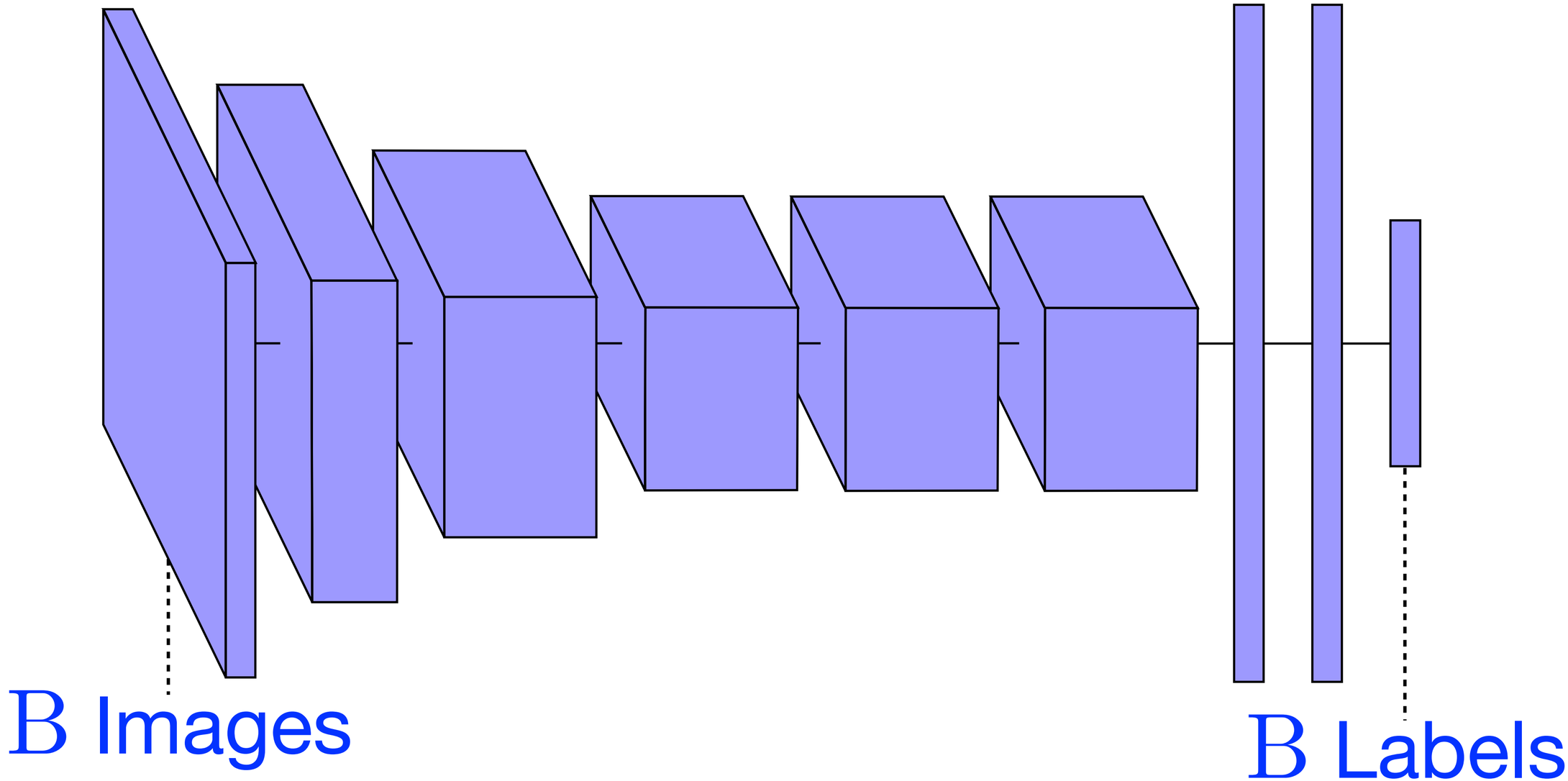

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
transfer AnB

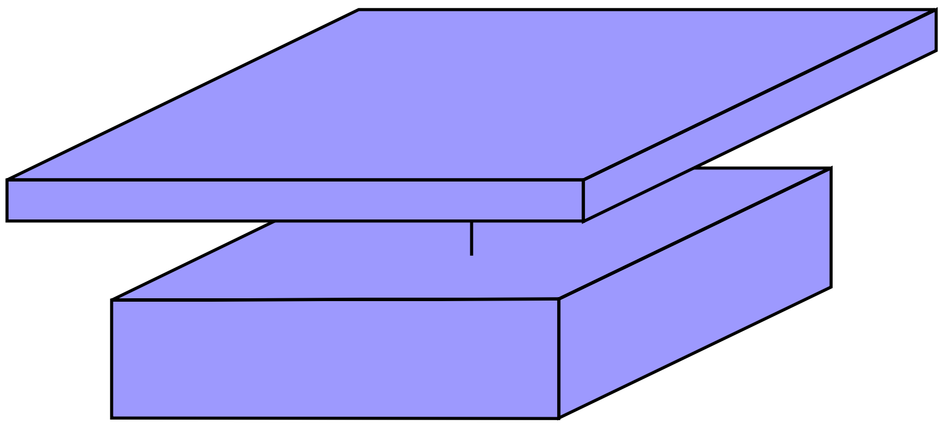

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
transfer AnB

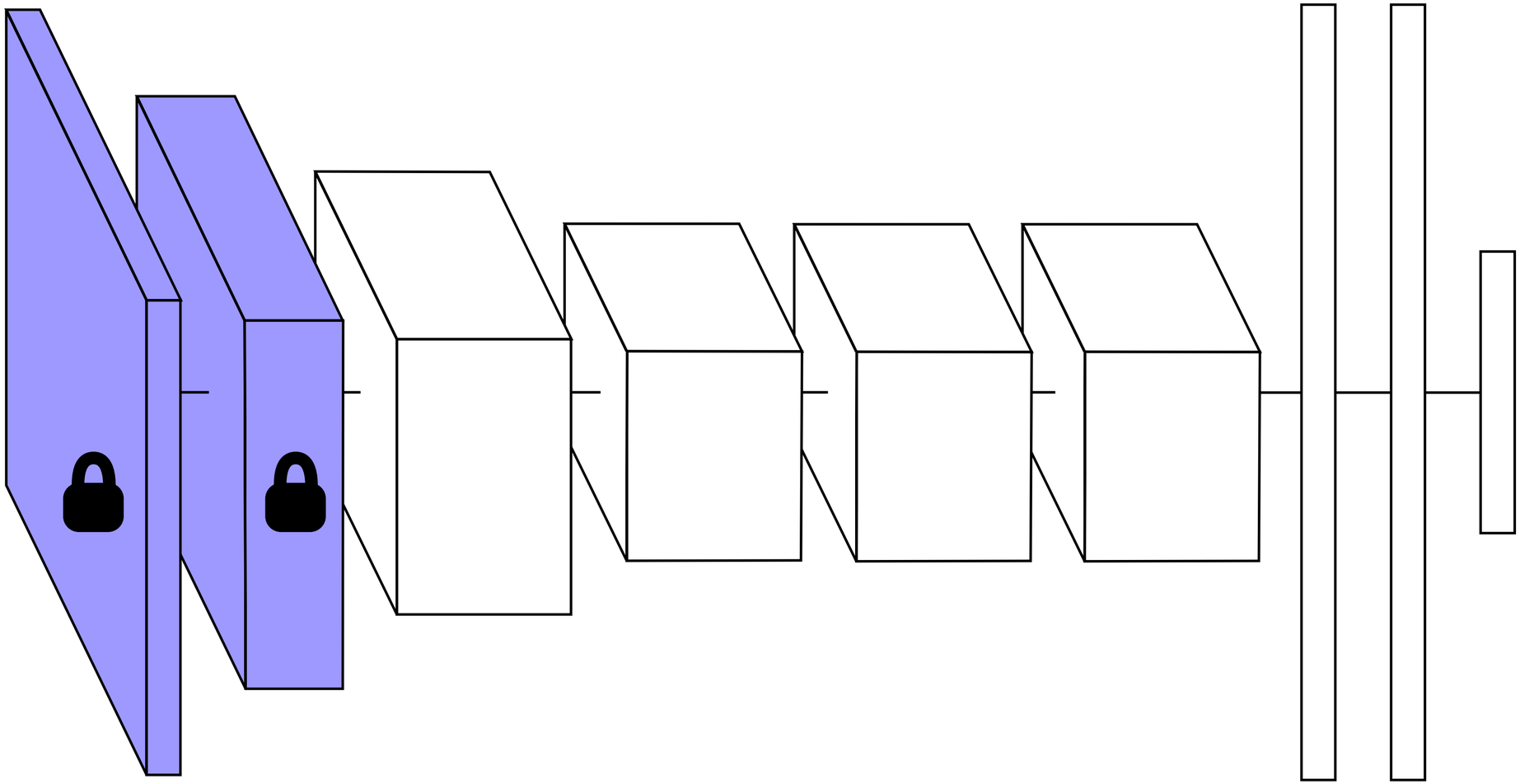

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
transfer AnB

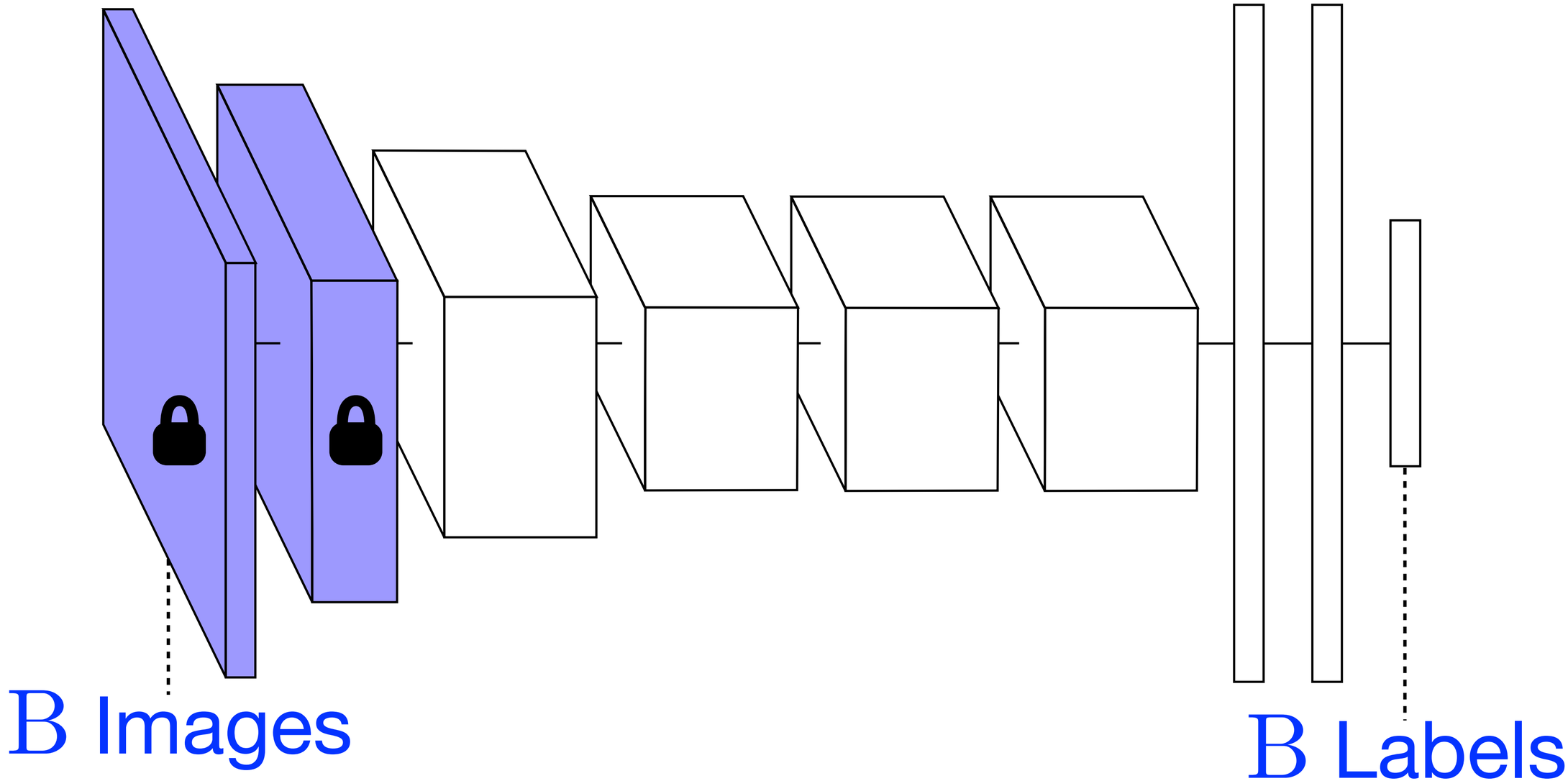

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
transfer AnB

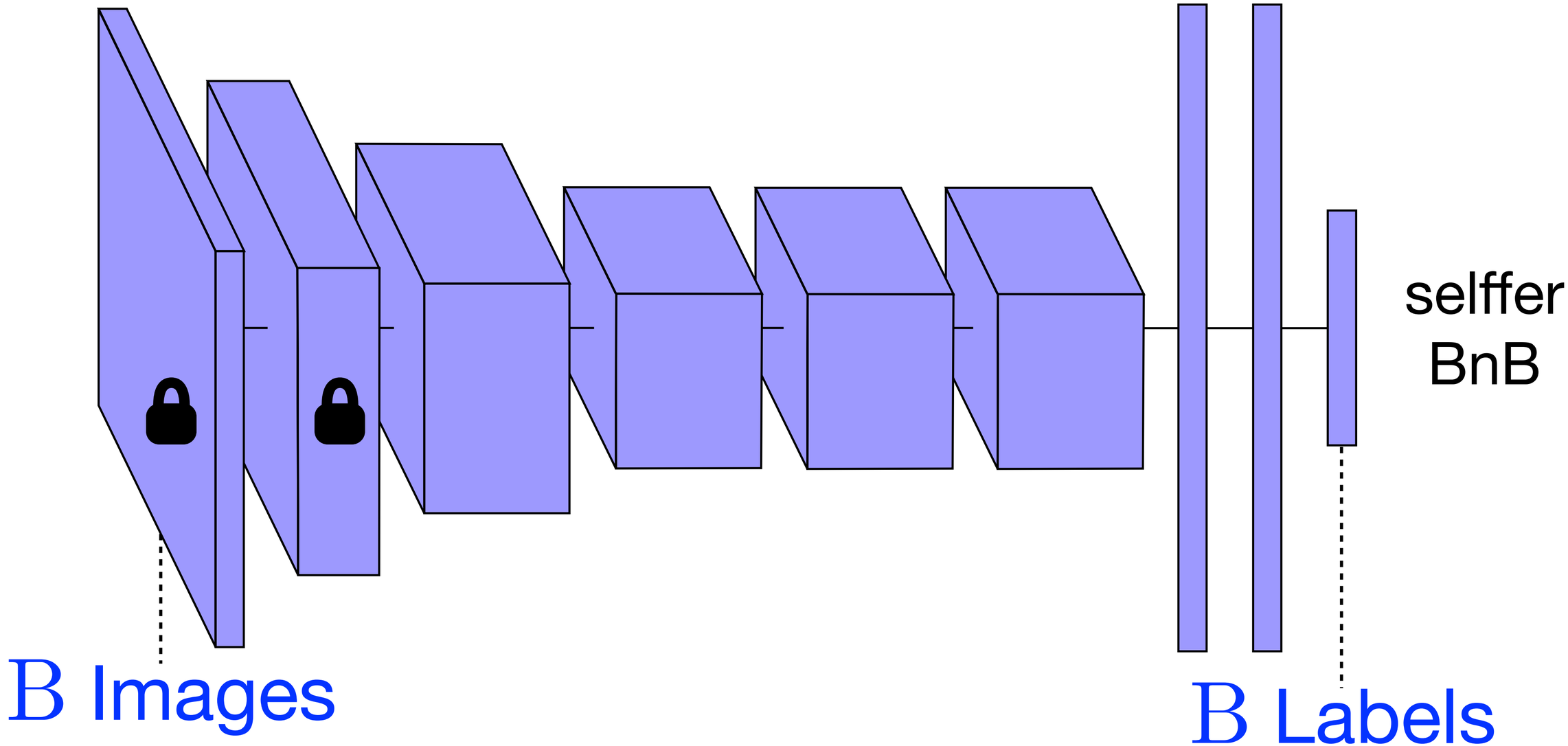

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
transfer AnB

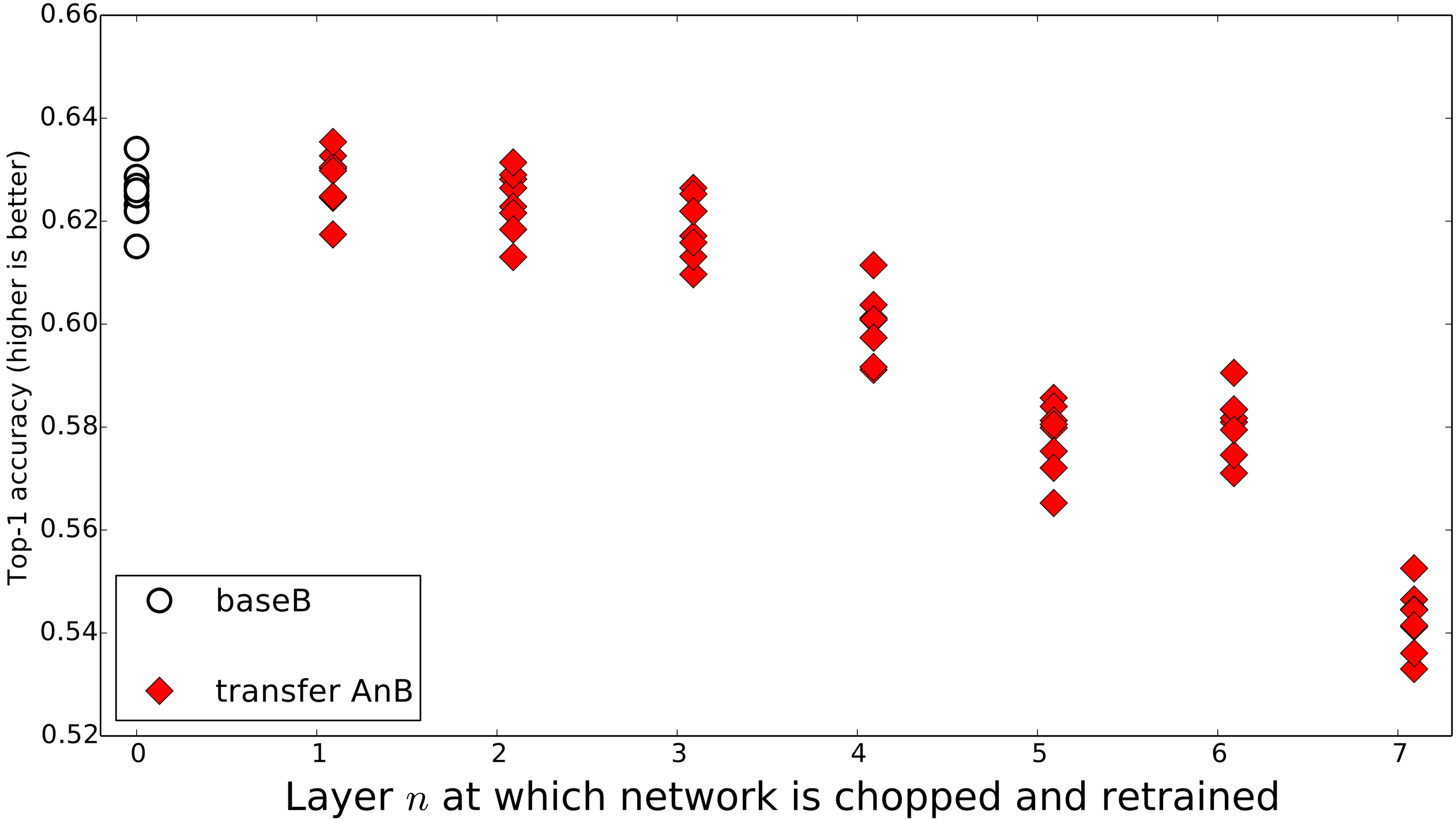

B Images
B Labels

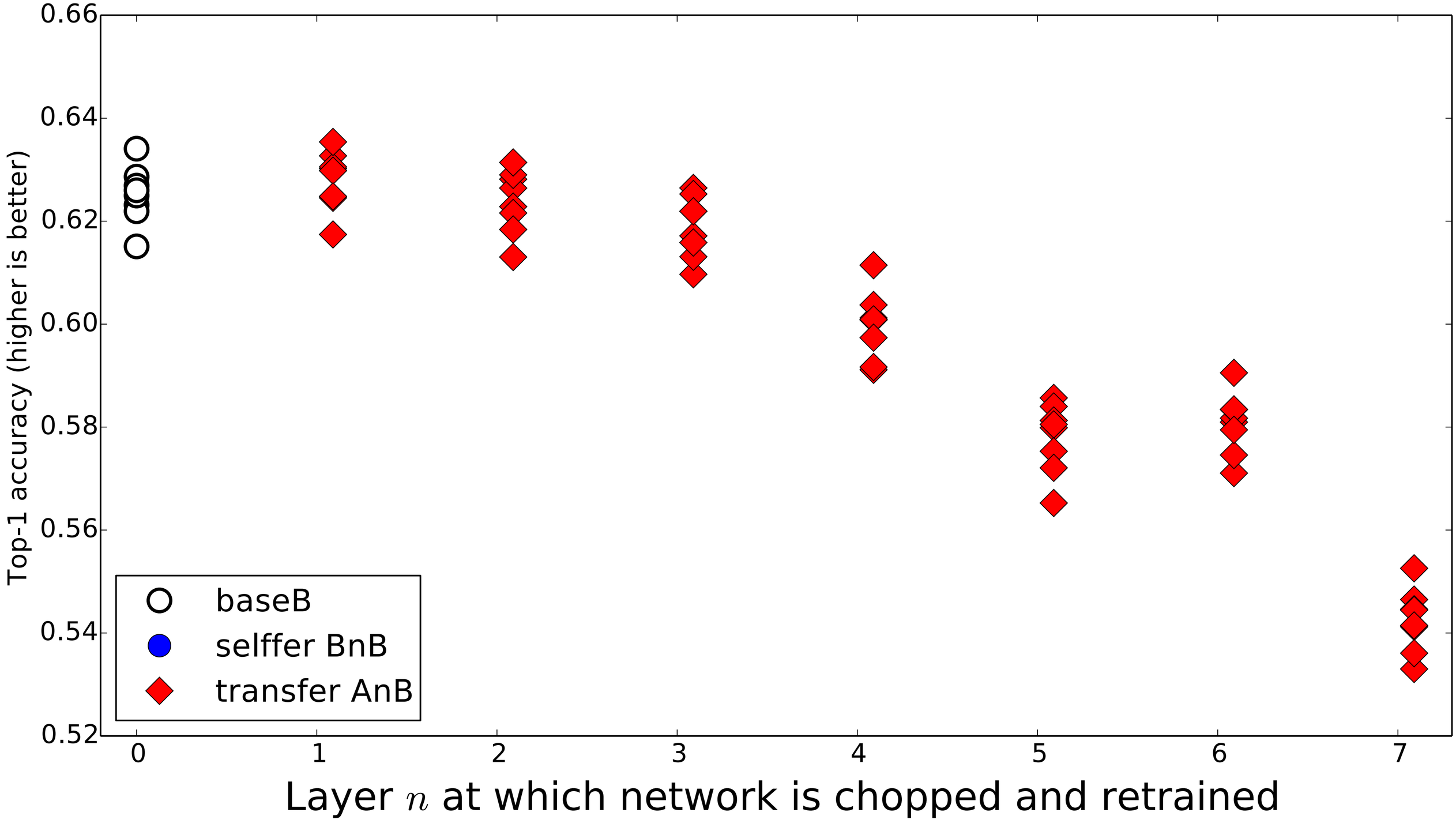


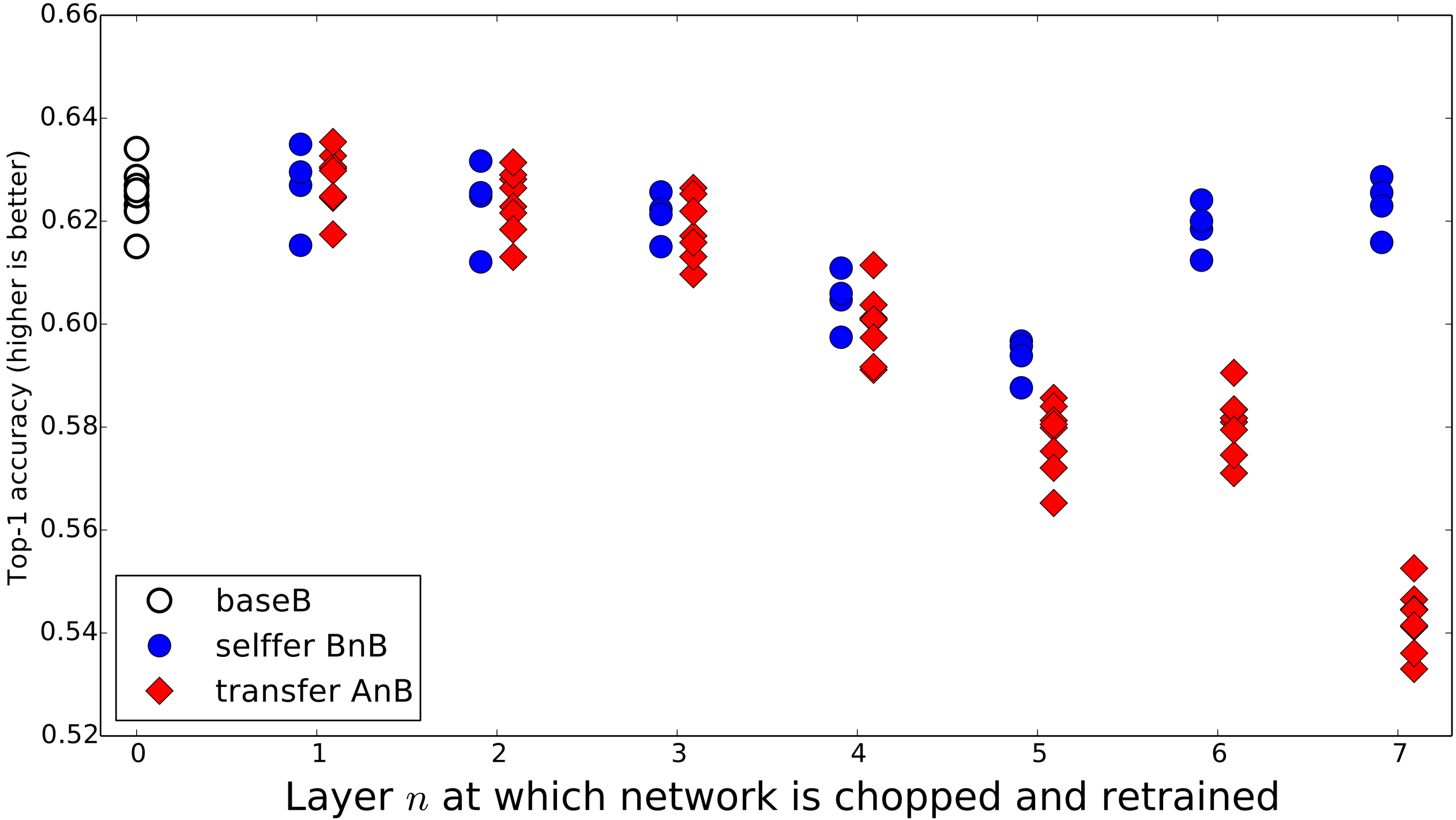


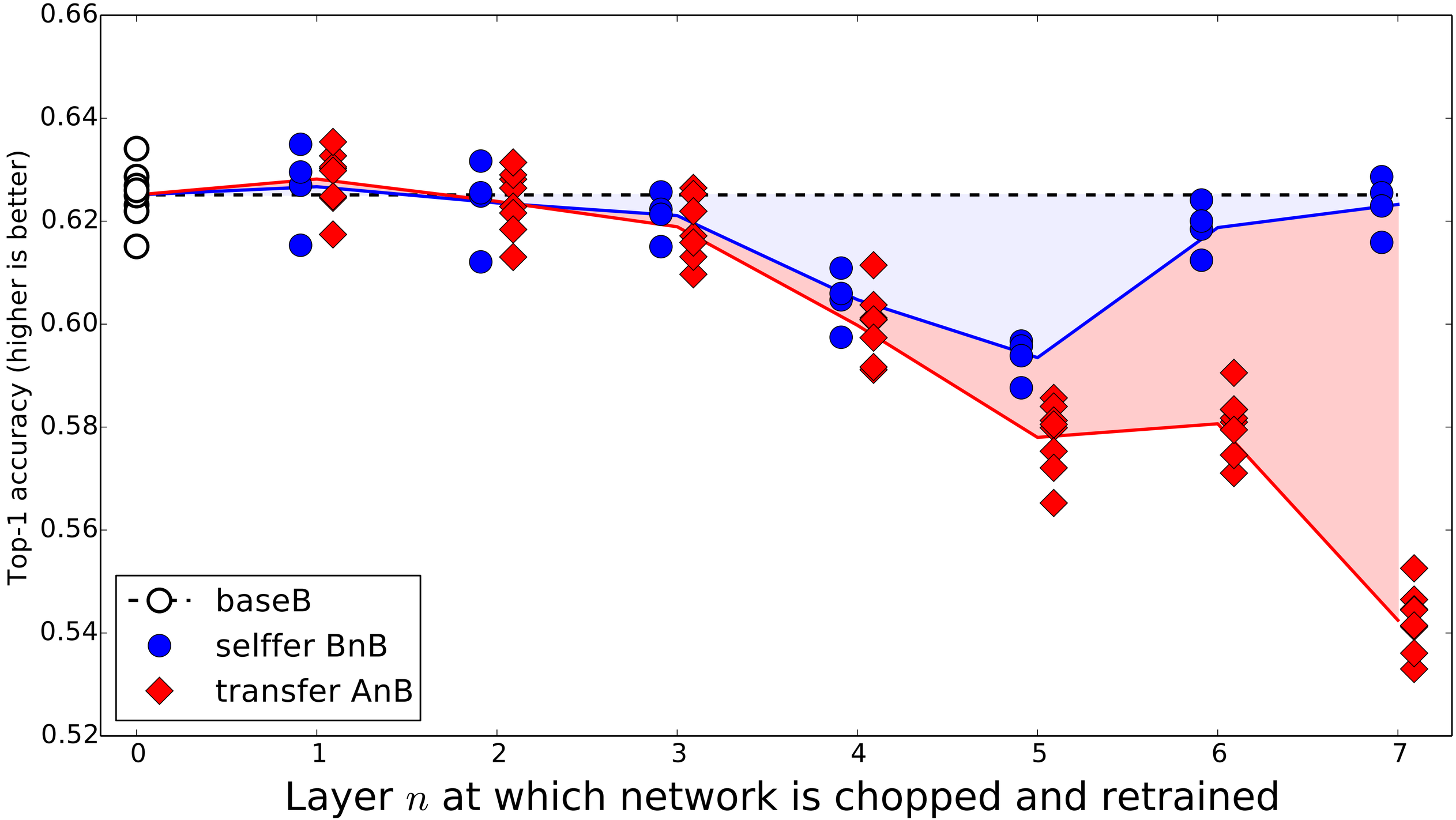

B Images
B Labels

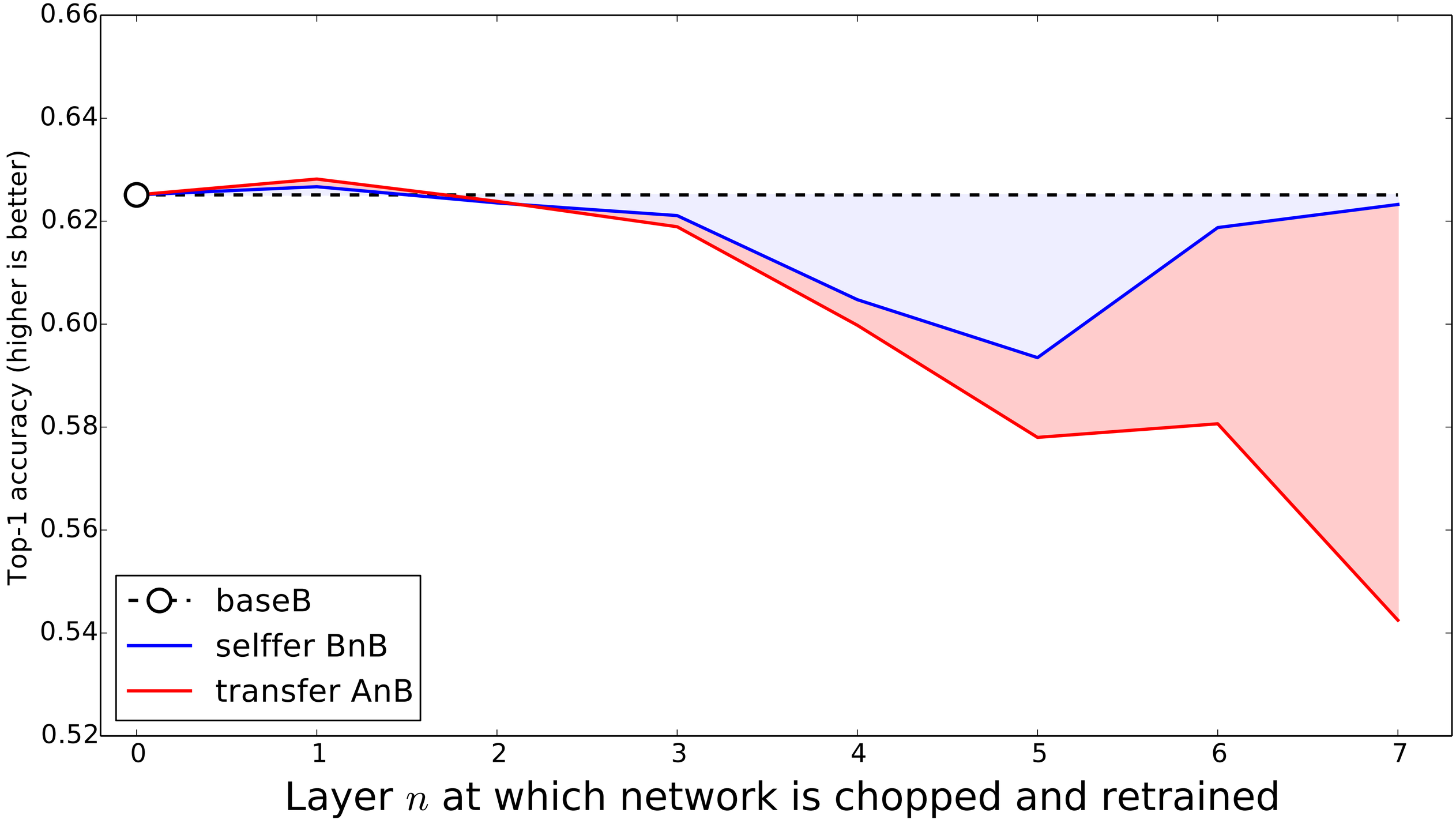

selffer
BnB
B Images
B Labels

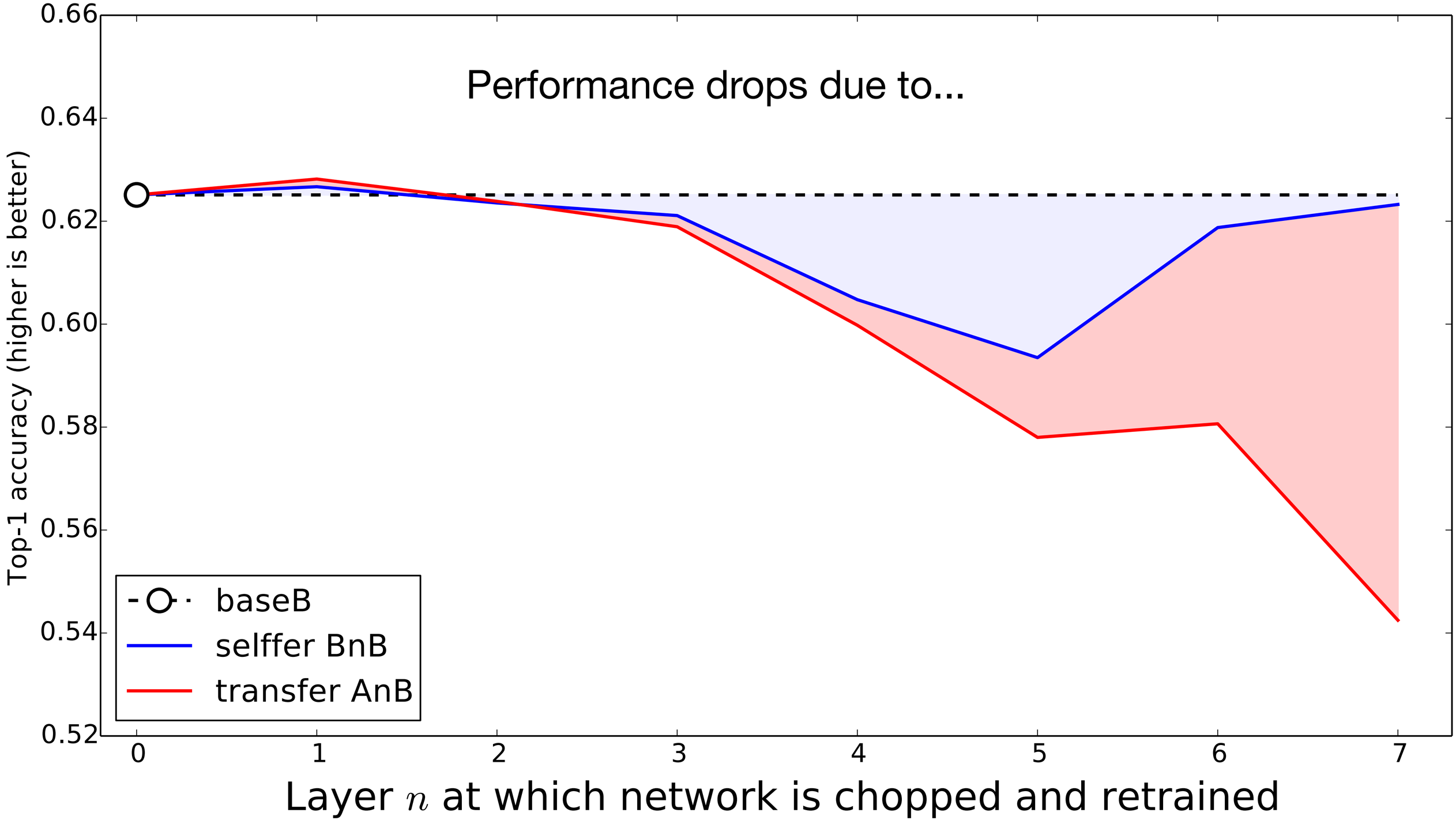

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
transfer AnB

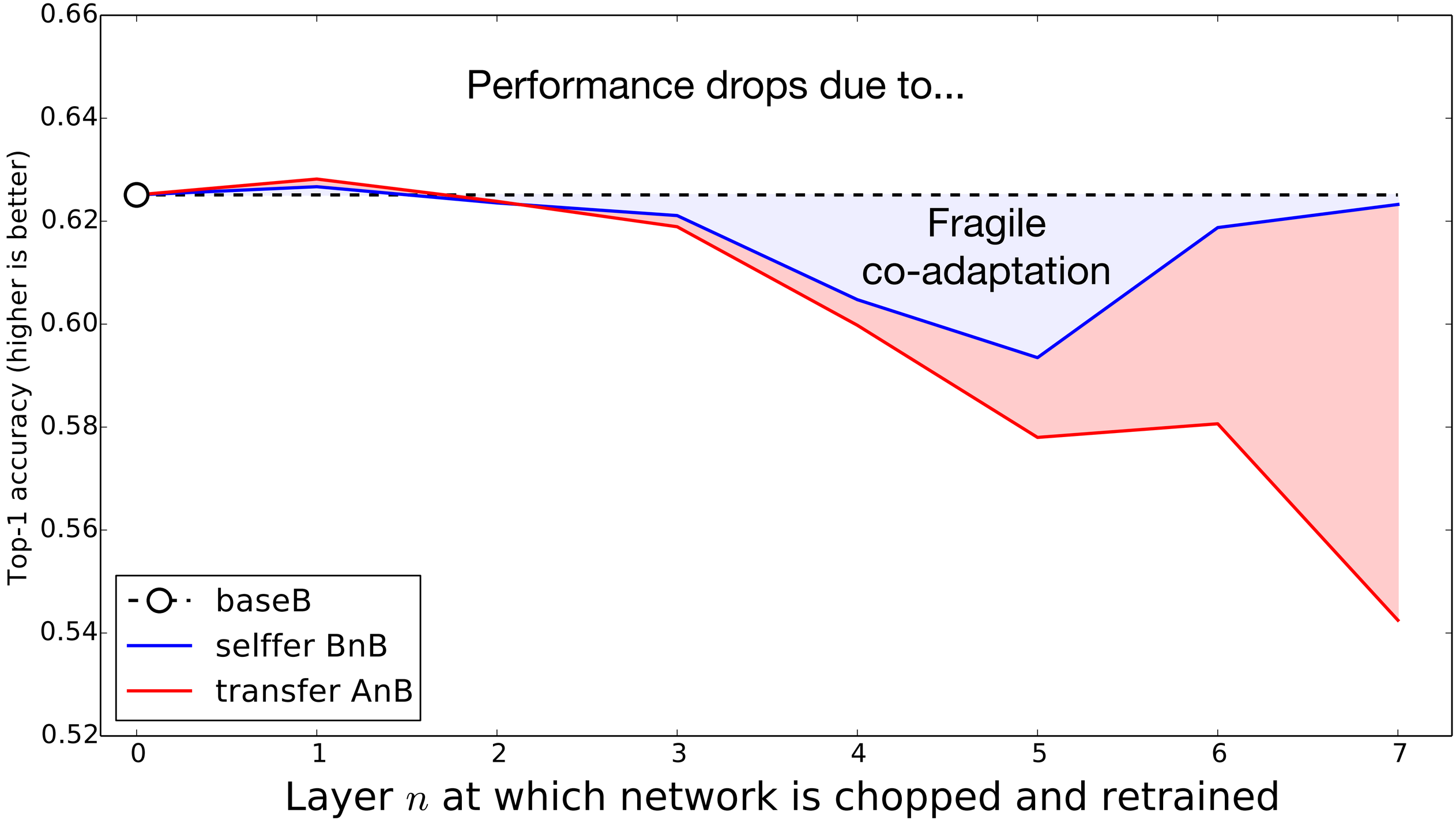

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
transfer AnB

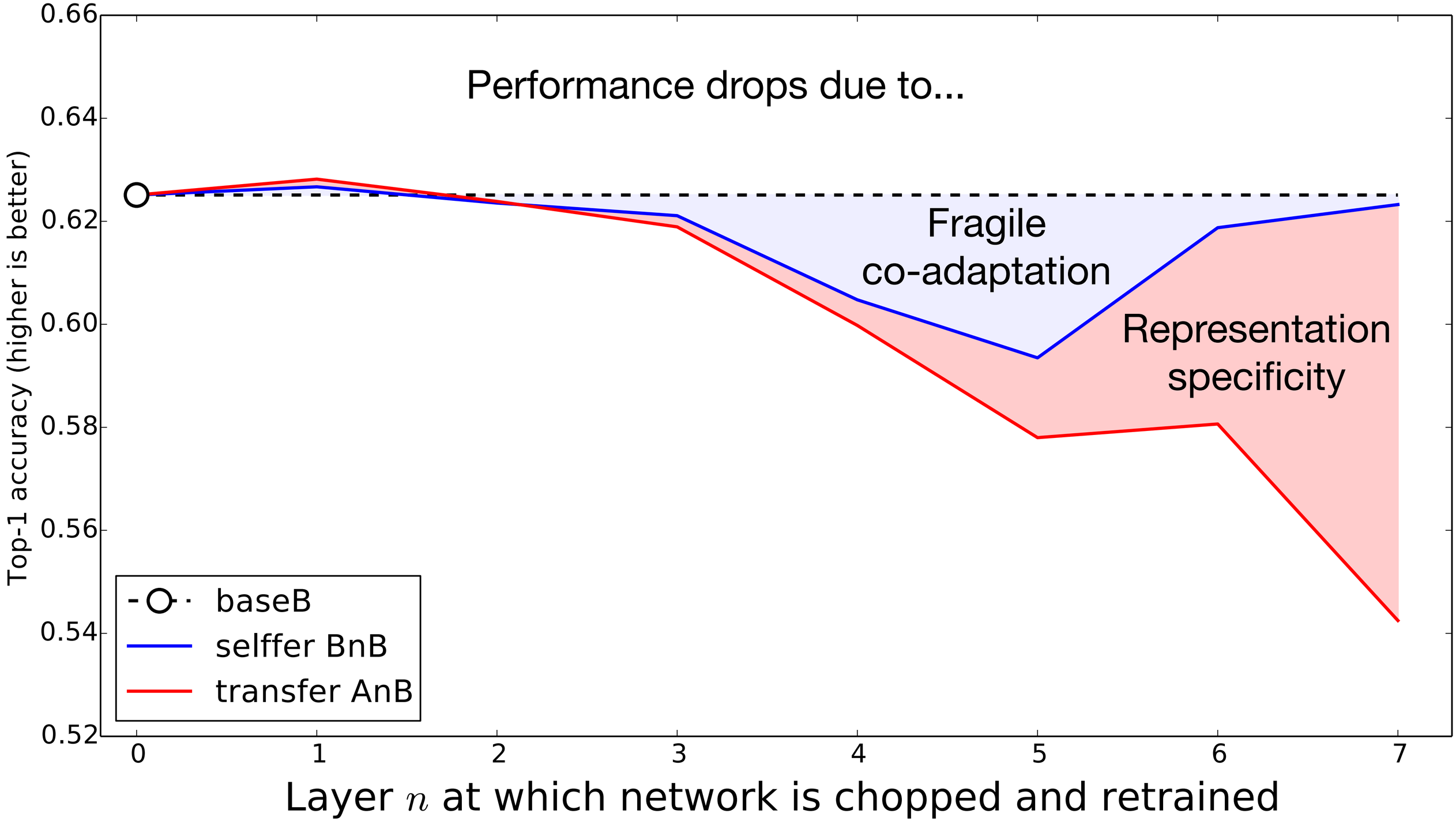

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
transfer AnB

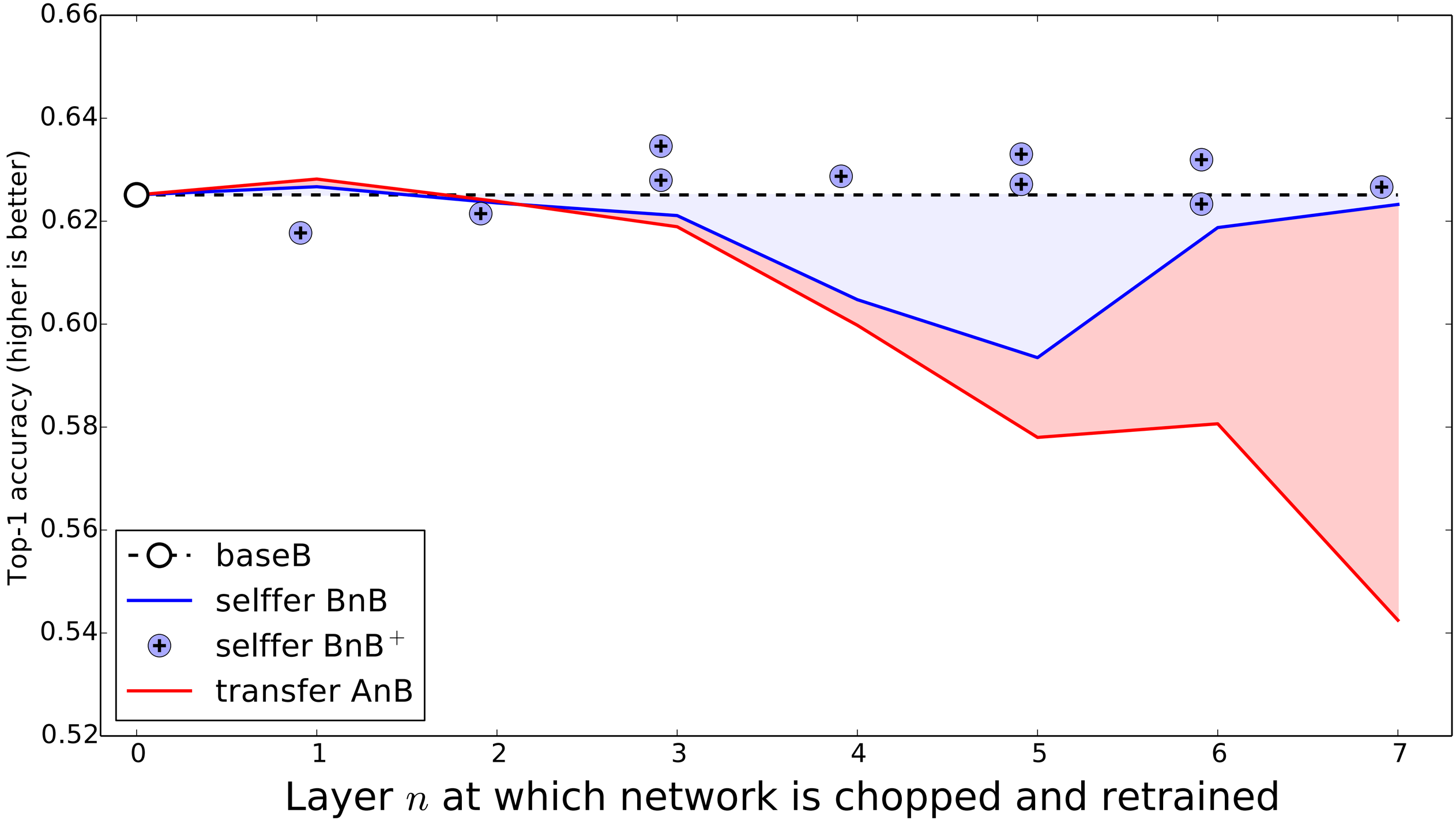

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
transfer AnB

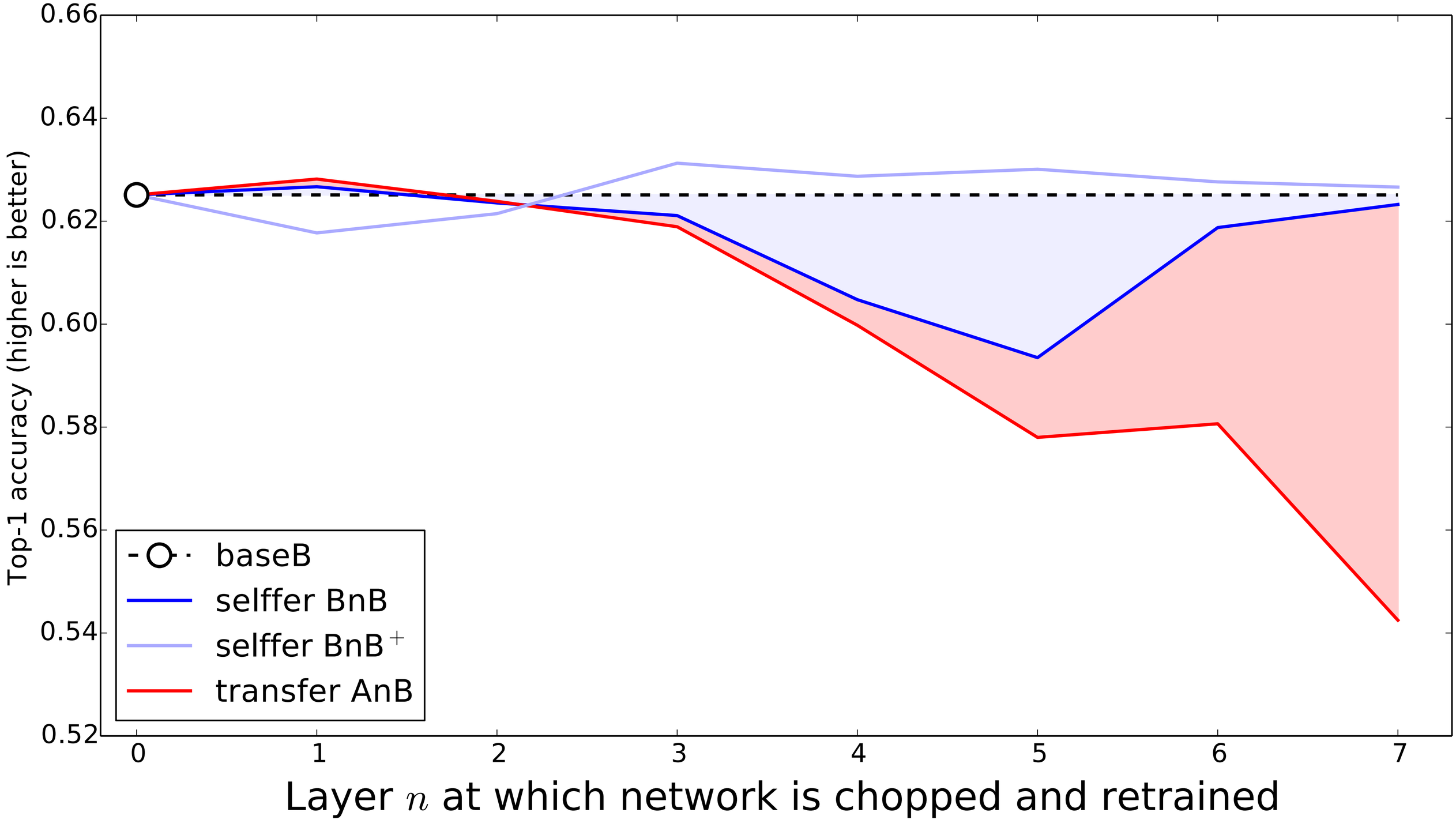

0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
transfer AnB

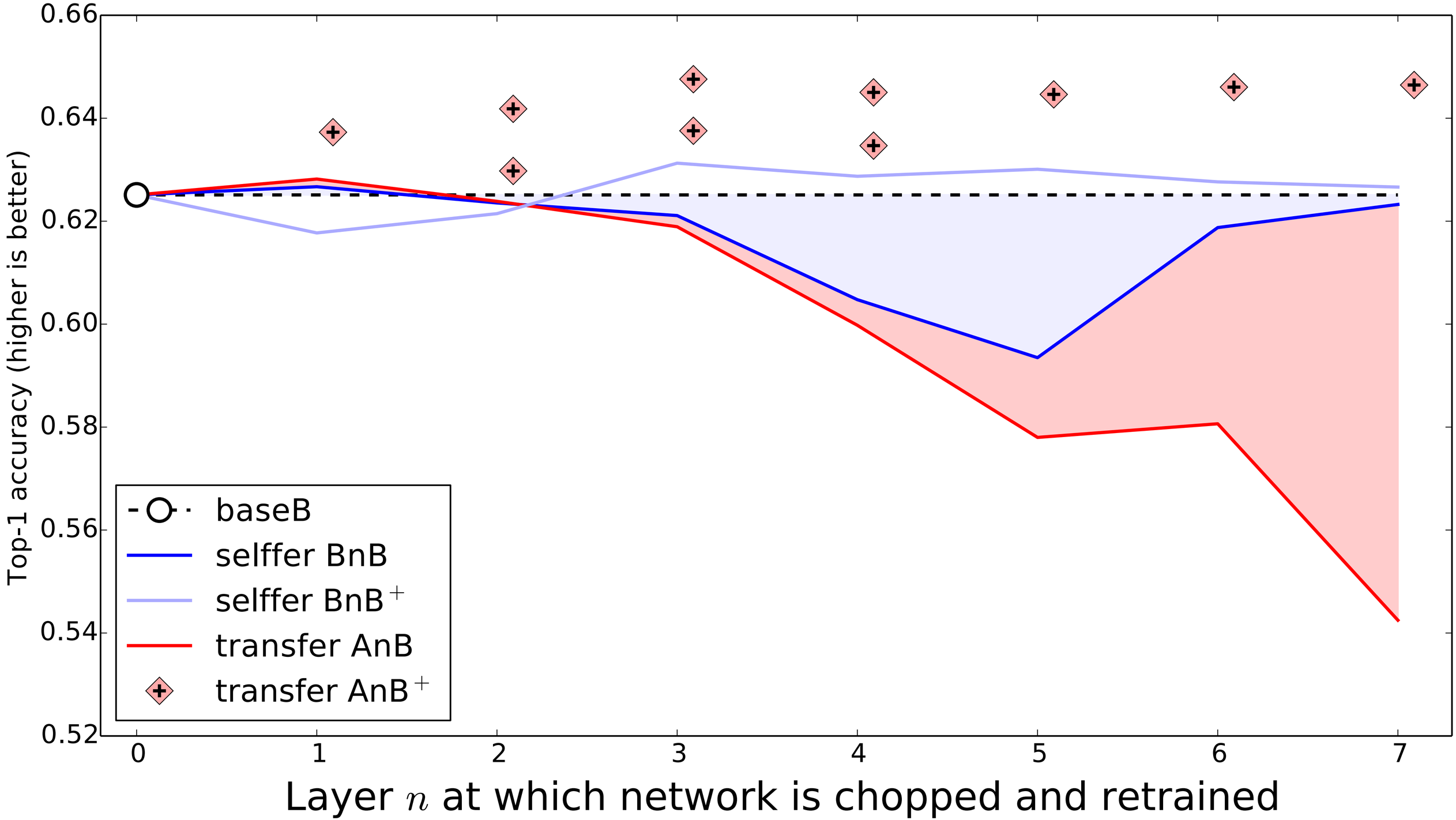

Performance drops due to...
Top-1 accuracy (higher is better)
0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
transfer AnB

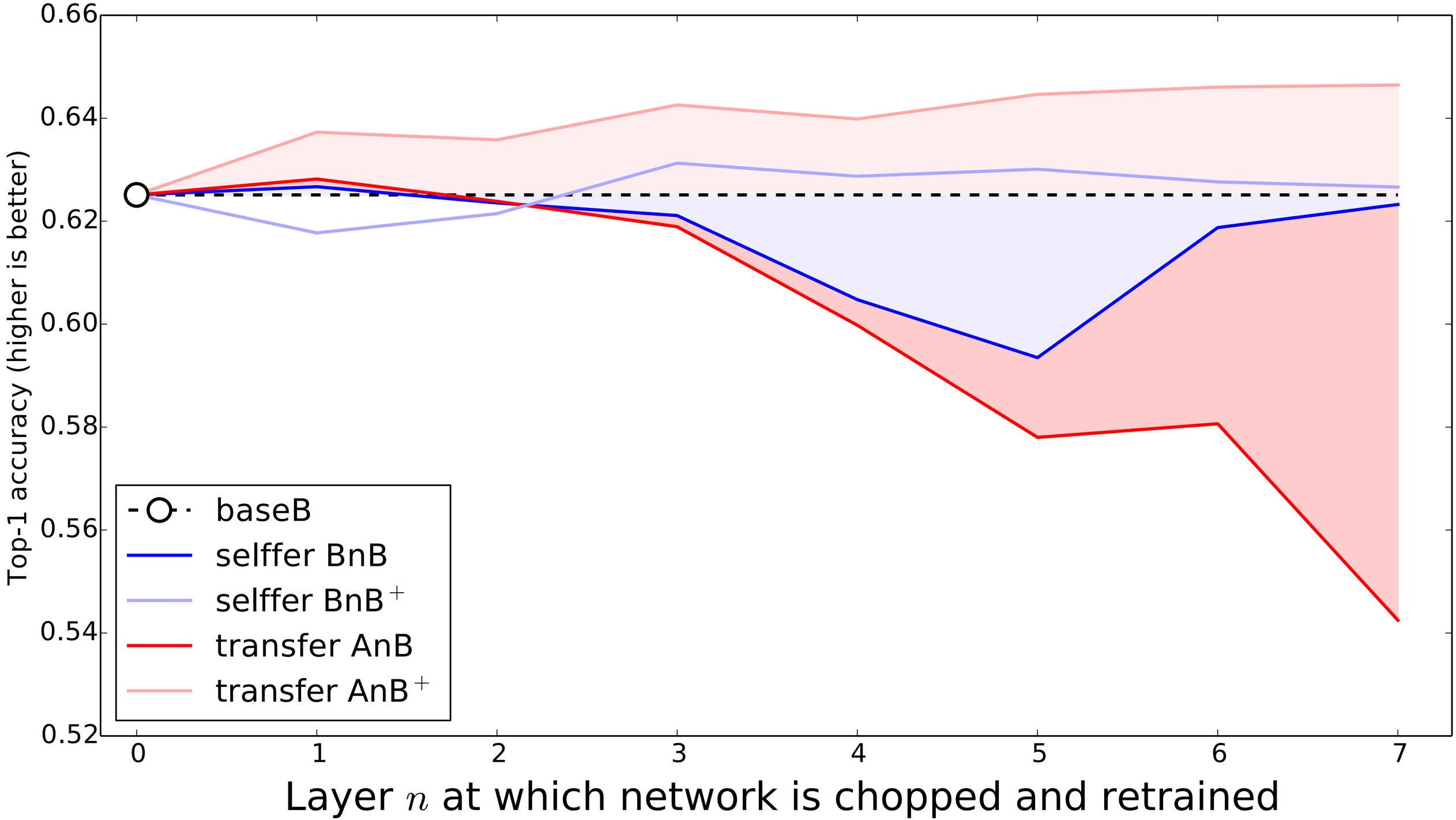

Performance drops due to...
Fragile
co-adaptation
Top-1 accuracy (higher is better)
0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
transfer AnB

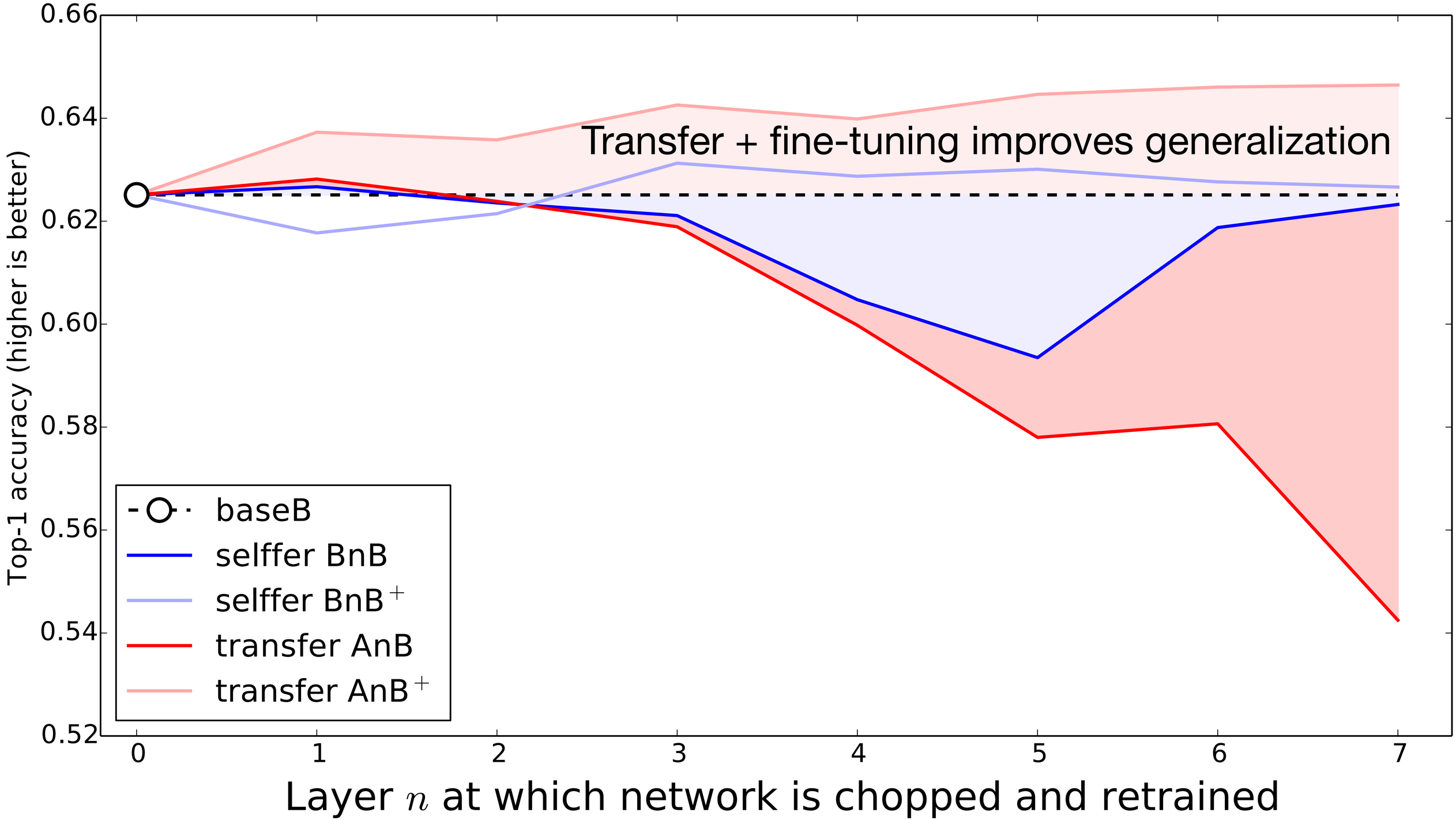

Performance drops due to...
Fragile
co-adaptation
Representation
specificity
baseB
selffer BnB
transfer AnB
Top-1 accuracy (higher is better)
0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained


0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
selffer BnB$^+$
transfer AnB


0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
selffer BnB$^+$
transfer AnB


0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
selffer BnB$^+$
transfer AnB
transfer AnB$^+$


0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
Top-1 accuracy (higher is better)
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
selffer BnB$^+$
transfer AnB
transfer AnB$^+$


Transfer + fine-tuning improves generalization
Top-1 accuracy (higher is better)
0.66
0.64
0.62
0.60
0.58
0.56
0.54
0.52
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
baseB
selffer BnB
selffer BnB$^+$
transfer AnB
transfer AnB$^+$

ImageNet has many related categories...

Dataset A: random

gecko

fire truck

baseball

panther

rabbit

gorilla

Dataset B: random

garbage truck

toucan

radiator

binoculars

lion

bookshop

ImageNet has many related categories...

Dataset A: man-made

garbage truck

fire truck

radiator

baseball

binoculars

bookshop

Dataset B: natural

gorilla

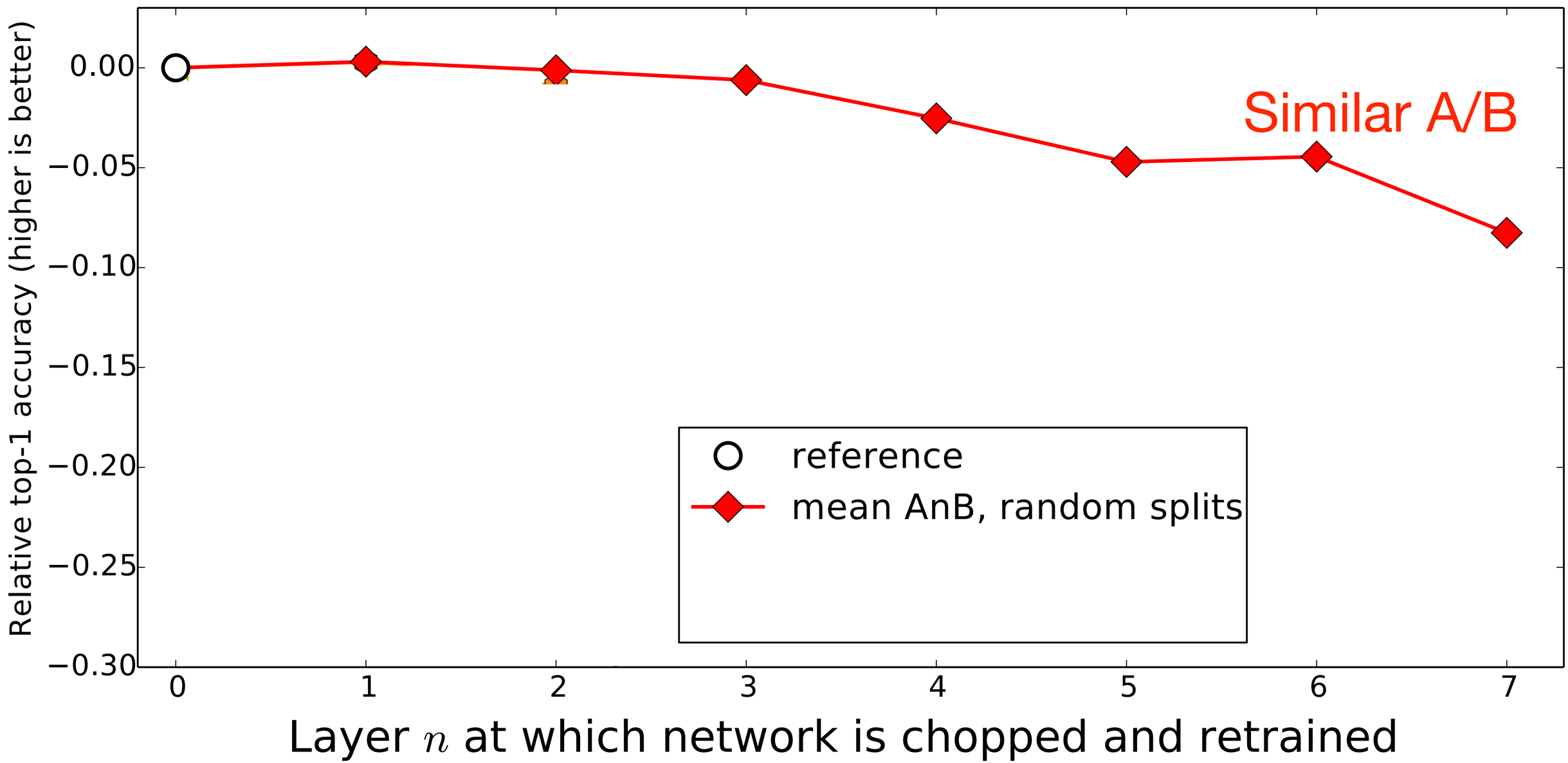
gecko

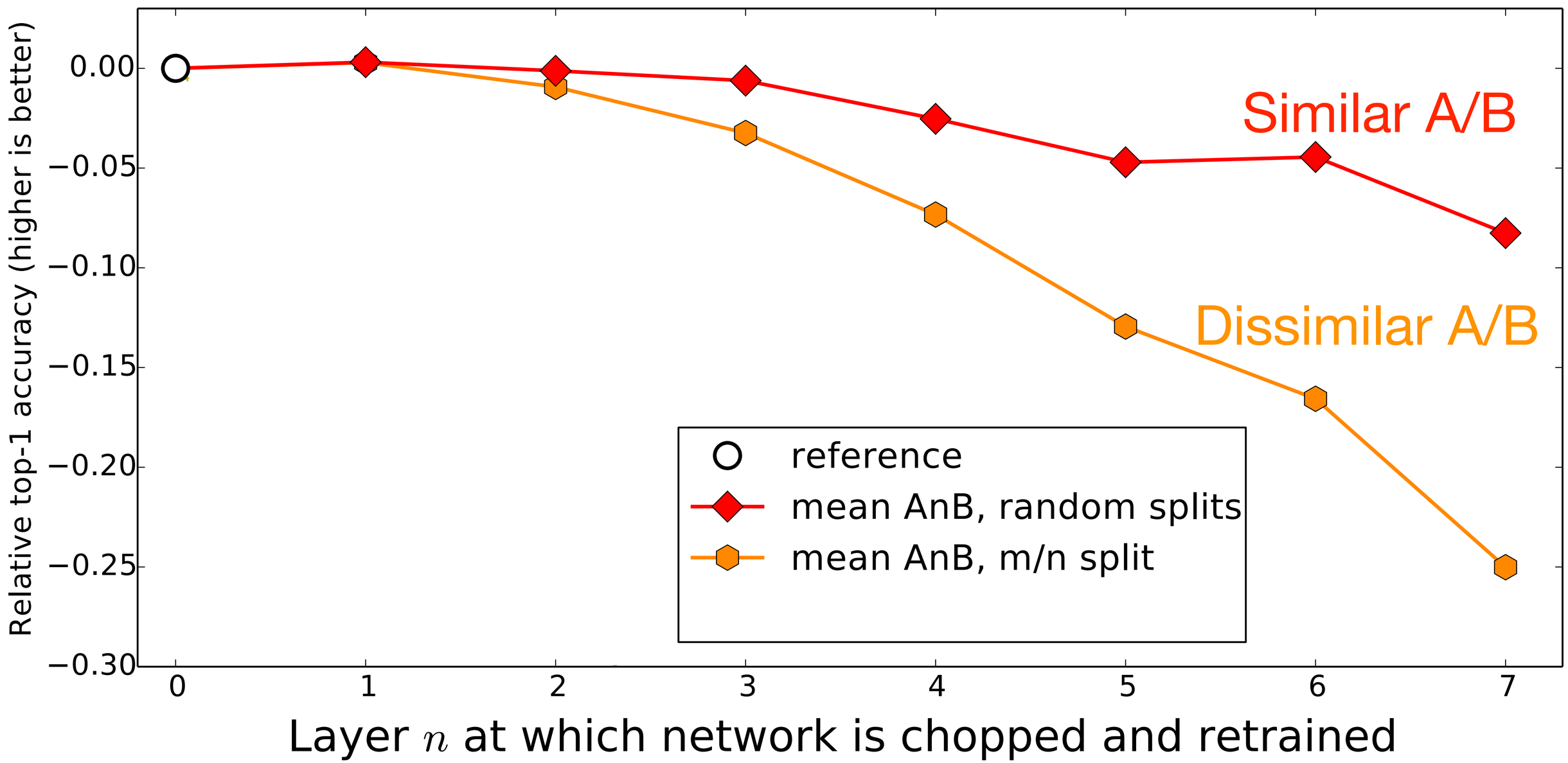
toucan

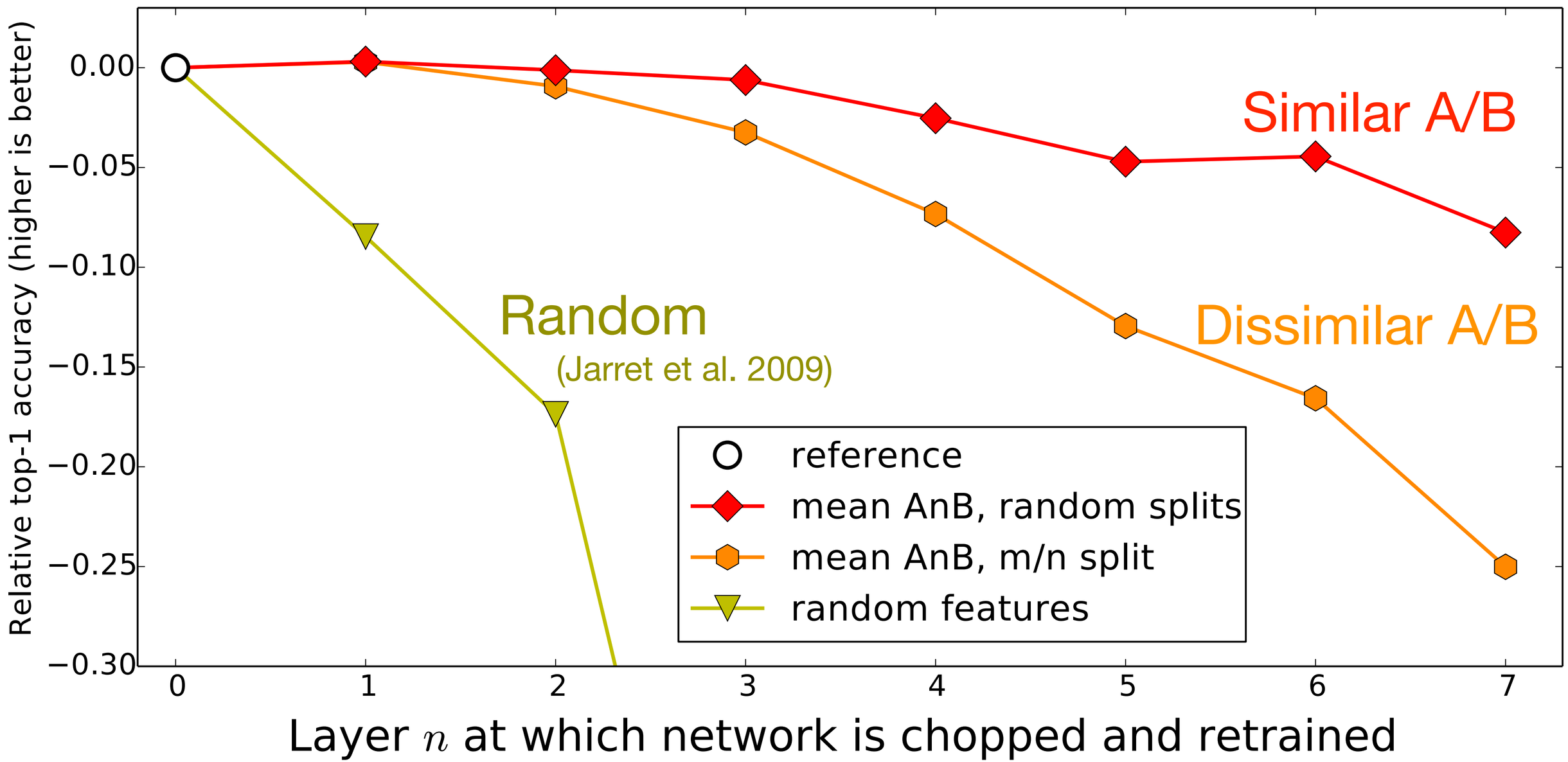
rabbit

panther

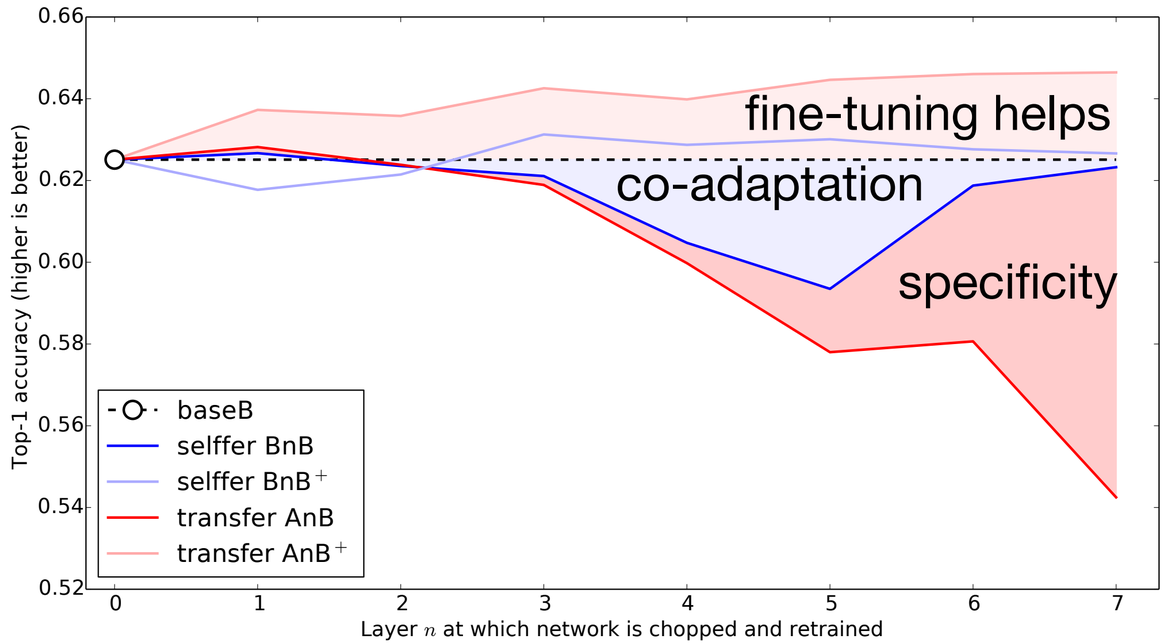
lion

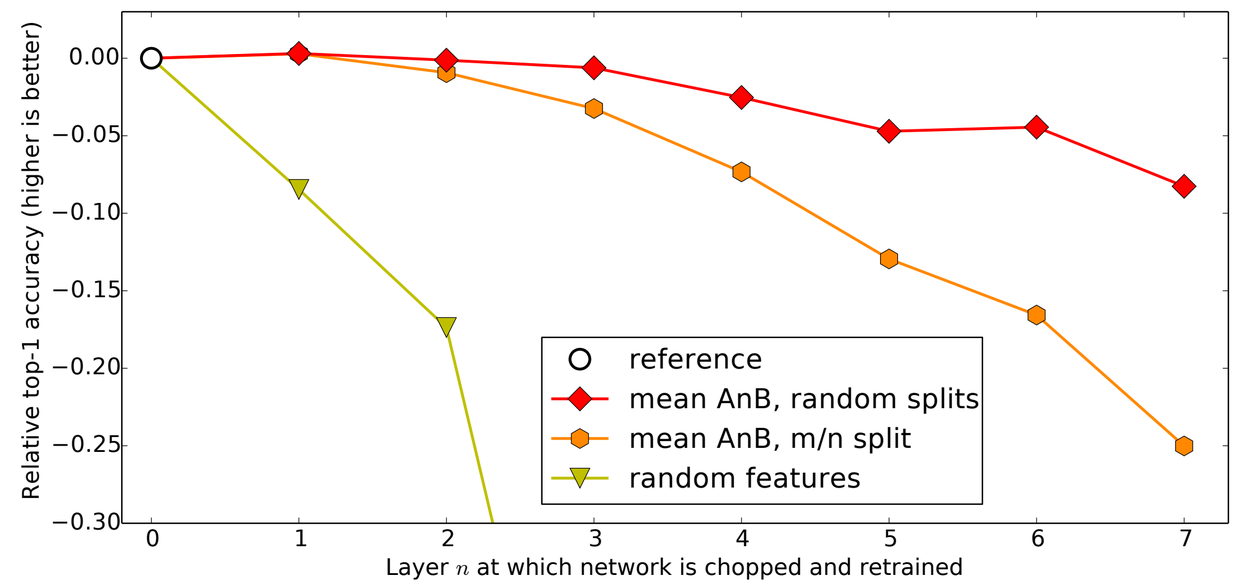

Similar A/B
reference
mean AnB, random splits
Relative top-1 accuracy (higher is better)
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.30
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained


Relative top-1 accuracy (higher is better)
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.30
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
Similar A/B
Dissimilar A/B
reference
mean AnB, random splits
mean AnB, m/n split

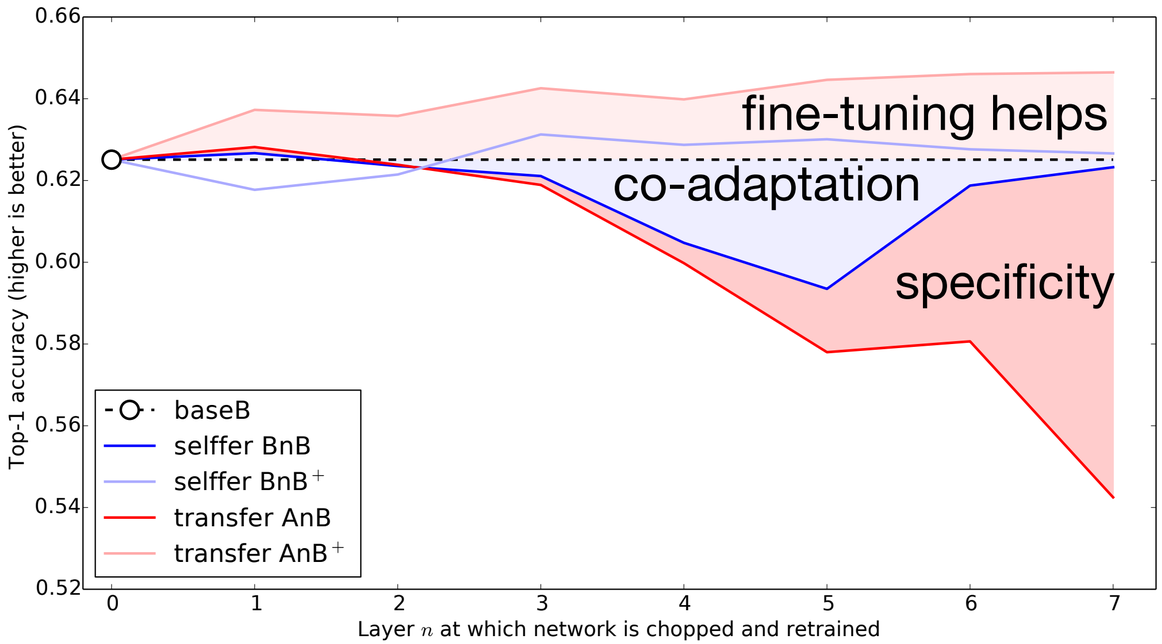

Relative top-1 accuracy (higher is better)
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.30
0
1
2
3
4
5
6
7
Layer n at which network is chopped and retrained
Similar A/B
Dissimilar A/B
Random
(Jarret et al. 2009)
reference
mean AnB, random splits
mean AnB, m/n split
random features

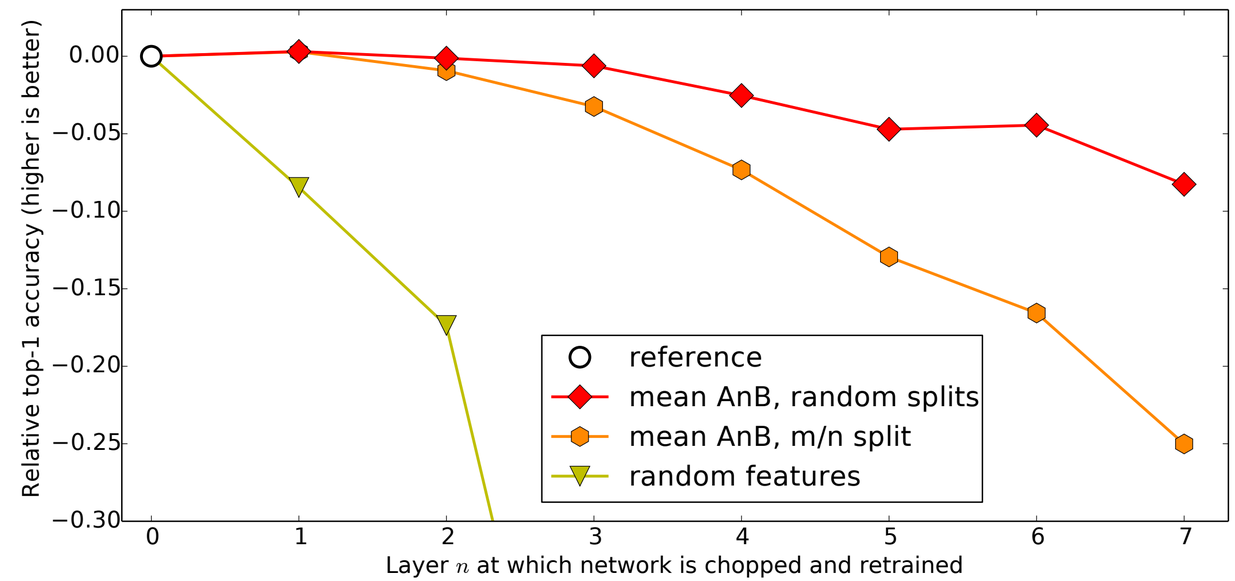
Conclusions





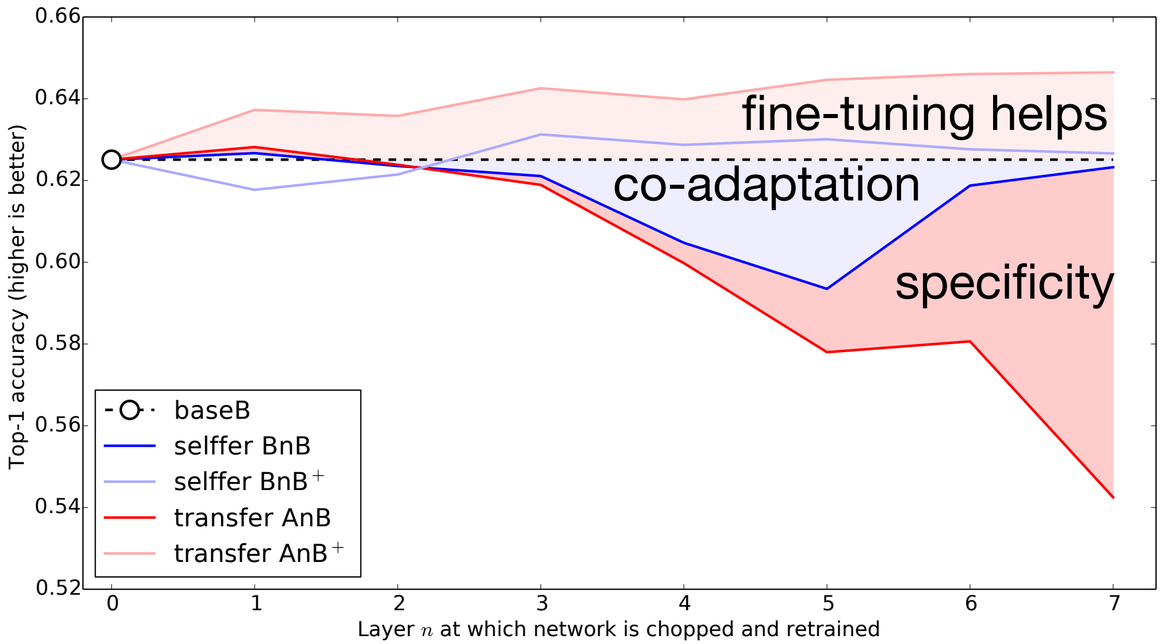
Conclusions

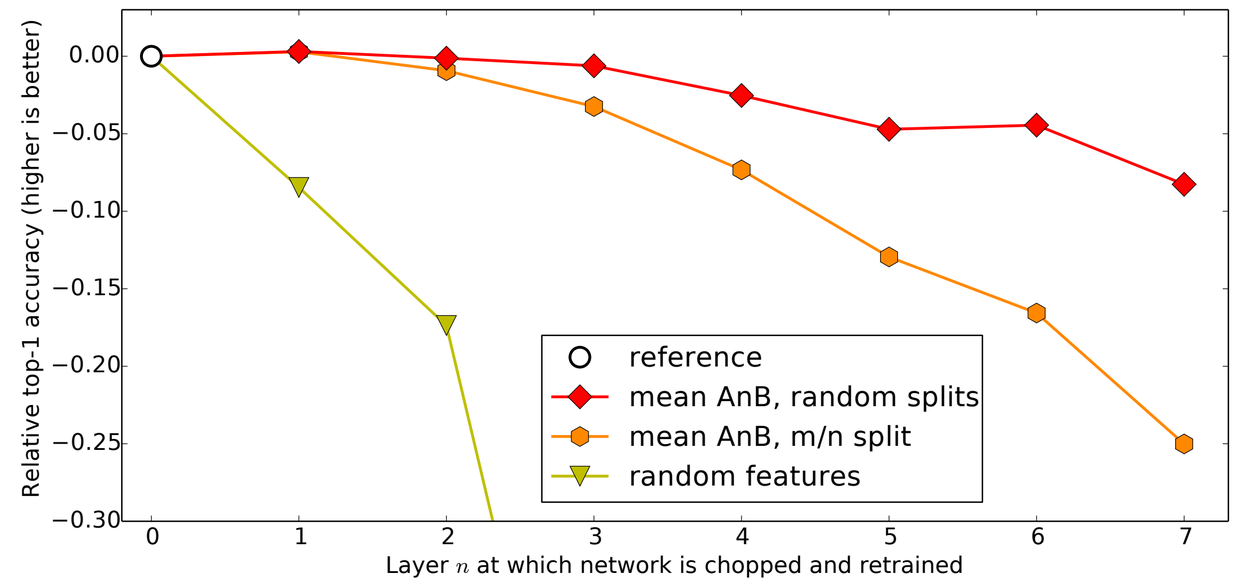




- Measure *general* to *specific* transition layer by layer

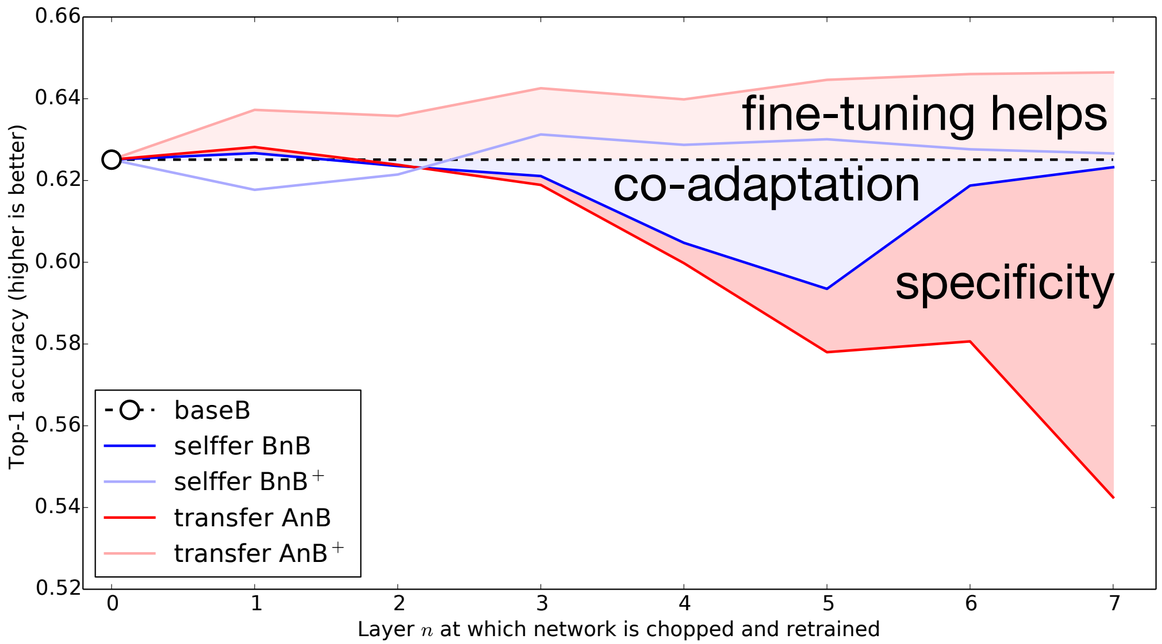
Conclusions

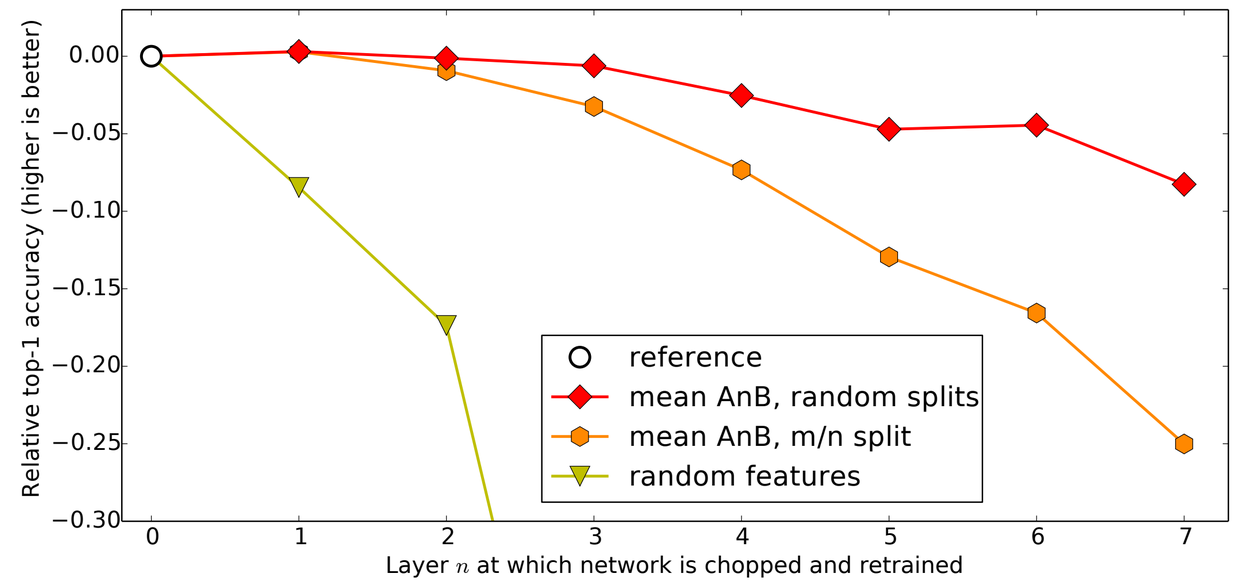




- Measure *general* to *specific* transition layer by layer
- Transferability governed by:

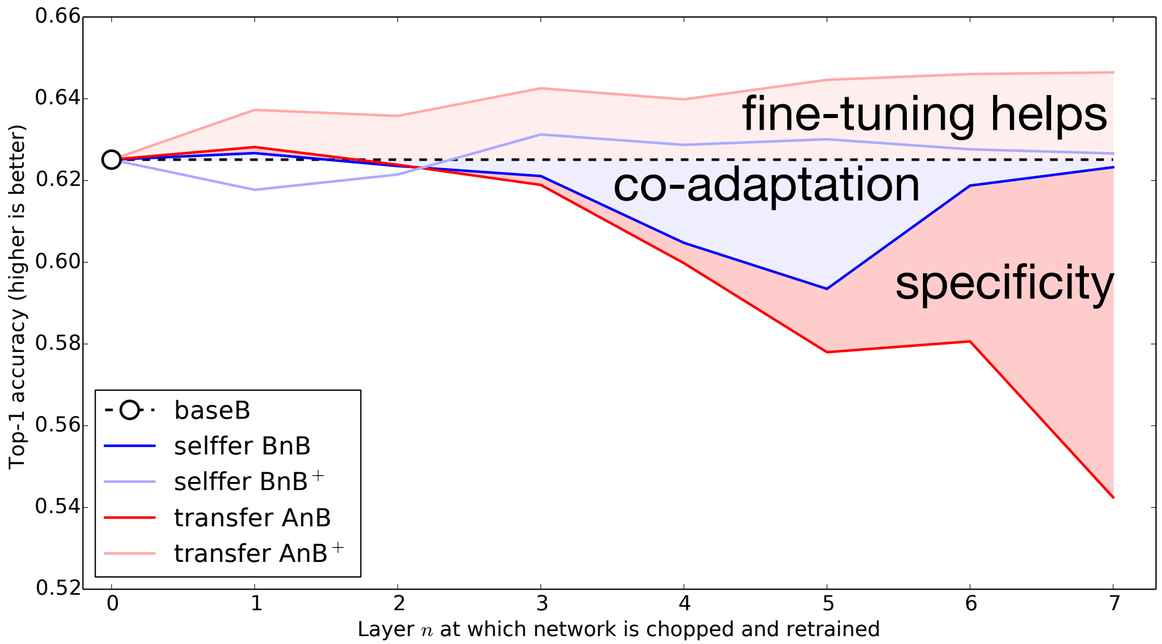
Conclusions

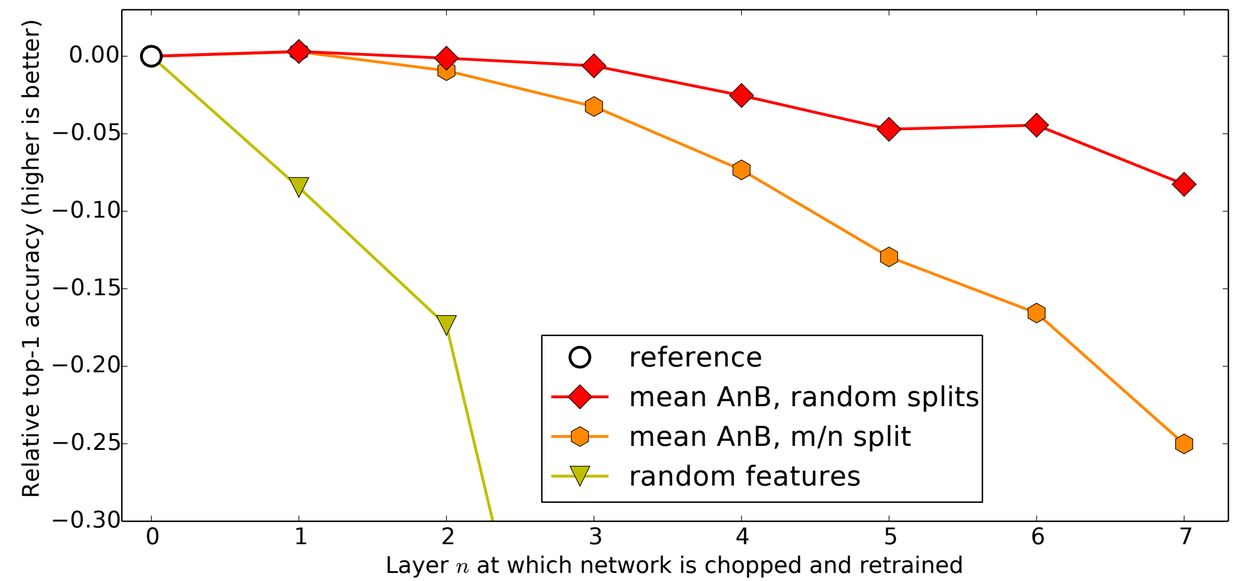




- Measure *general* to *specific* transition layer by layer
- Transferability governed by:
  - lost co-adaptations

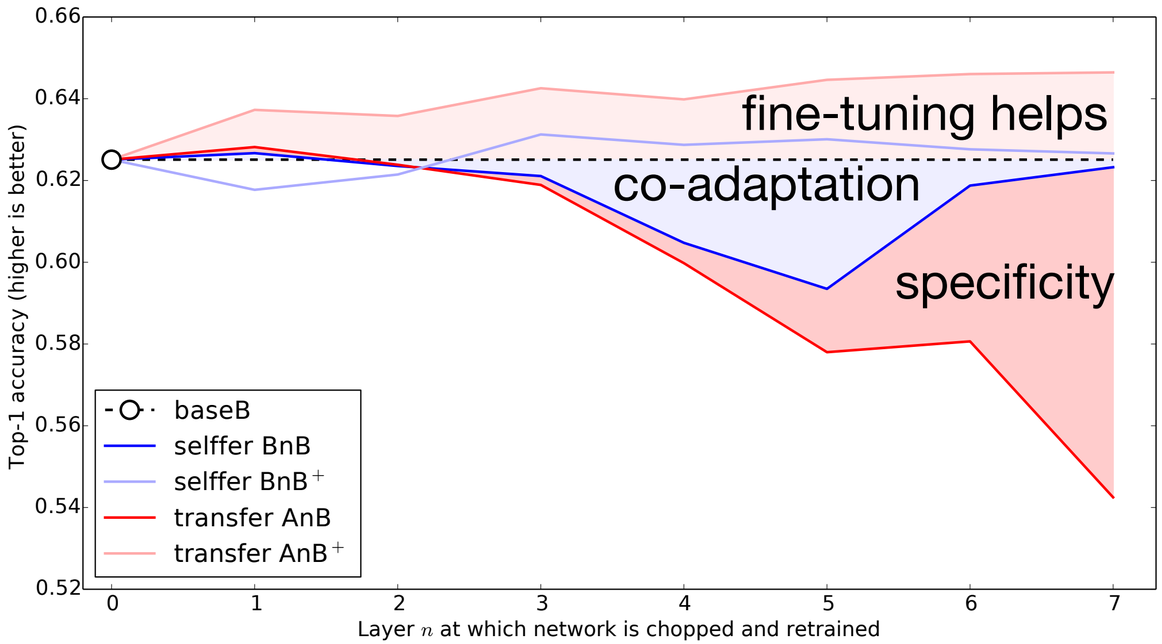
Conclusions

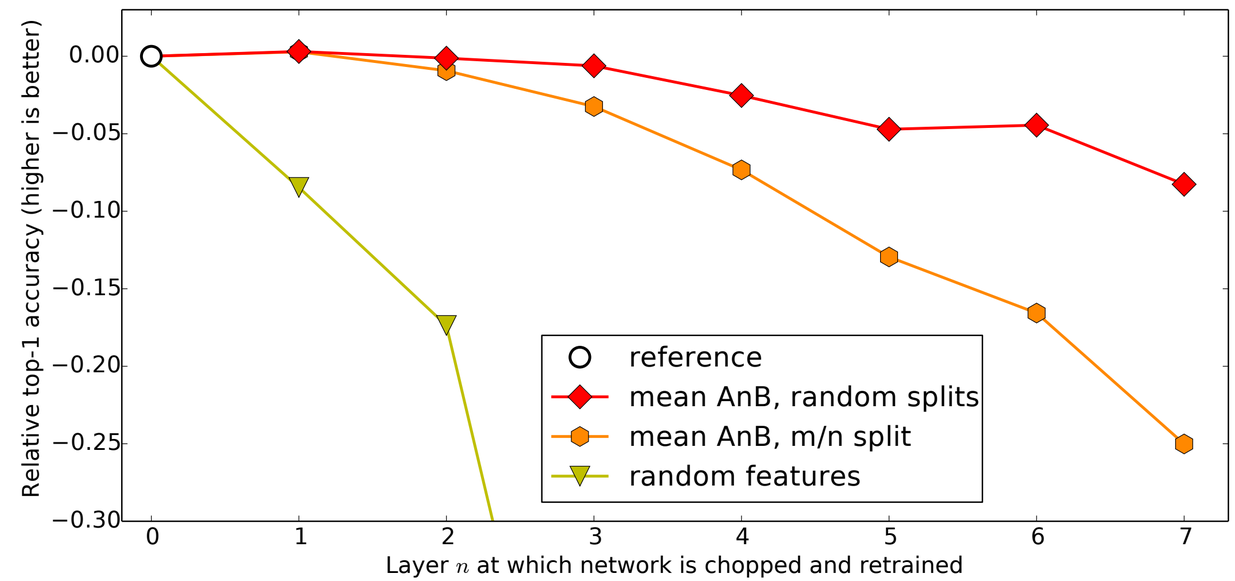




- Measure *general* to *specific* transition layer by layer
- Transferability governed by:
  - lost co-adaptations
  - specificity

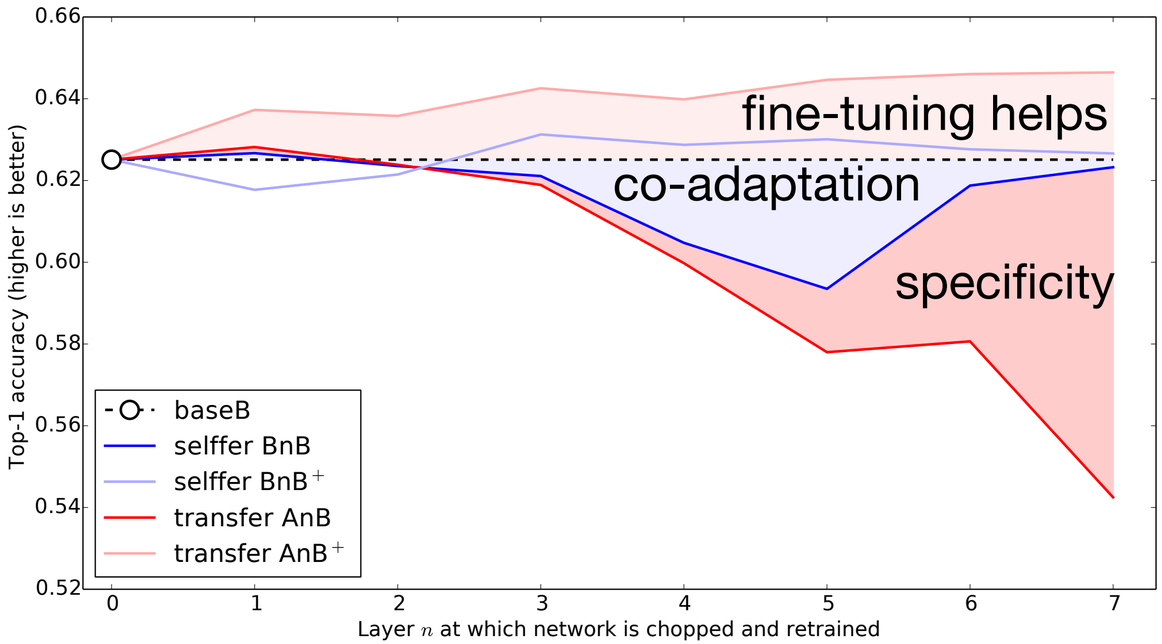
Conclusions

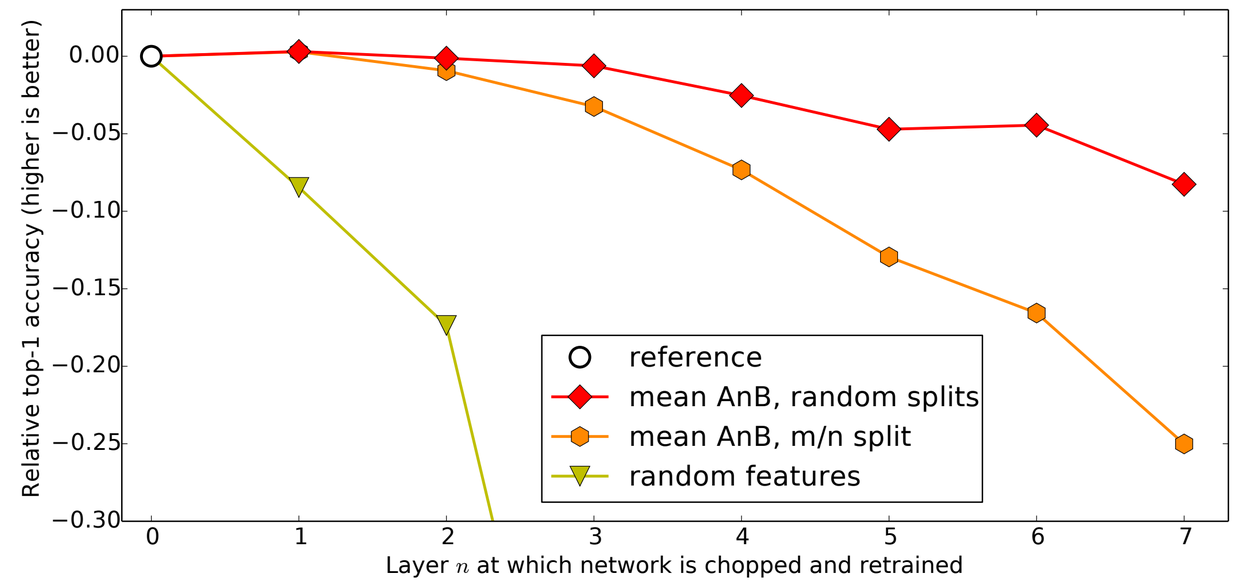




- Measure *general* to *specific* transition layer by layer
- Transferability governed by:
  - lost co-adaptations
  - specificity
  - difference between base and target dataset

Conclusions





- Measure *general* to *specific* transition layer by layer
- Transferability governed by:
  - lost co-adaptations
  - specificity
  - difference between base and target dataset
- Fine-tuning helps even on large target dataset