

Understanding Text on Images with AI at Scale

Viswanath Sivakumar

Facebook AI Research (FAIR)



PINGUINO FELIZ TE ESPERA

PINGUINO FELIZ TE ESPERA

EN MAGDALENA EN 2018

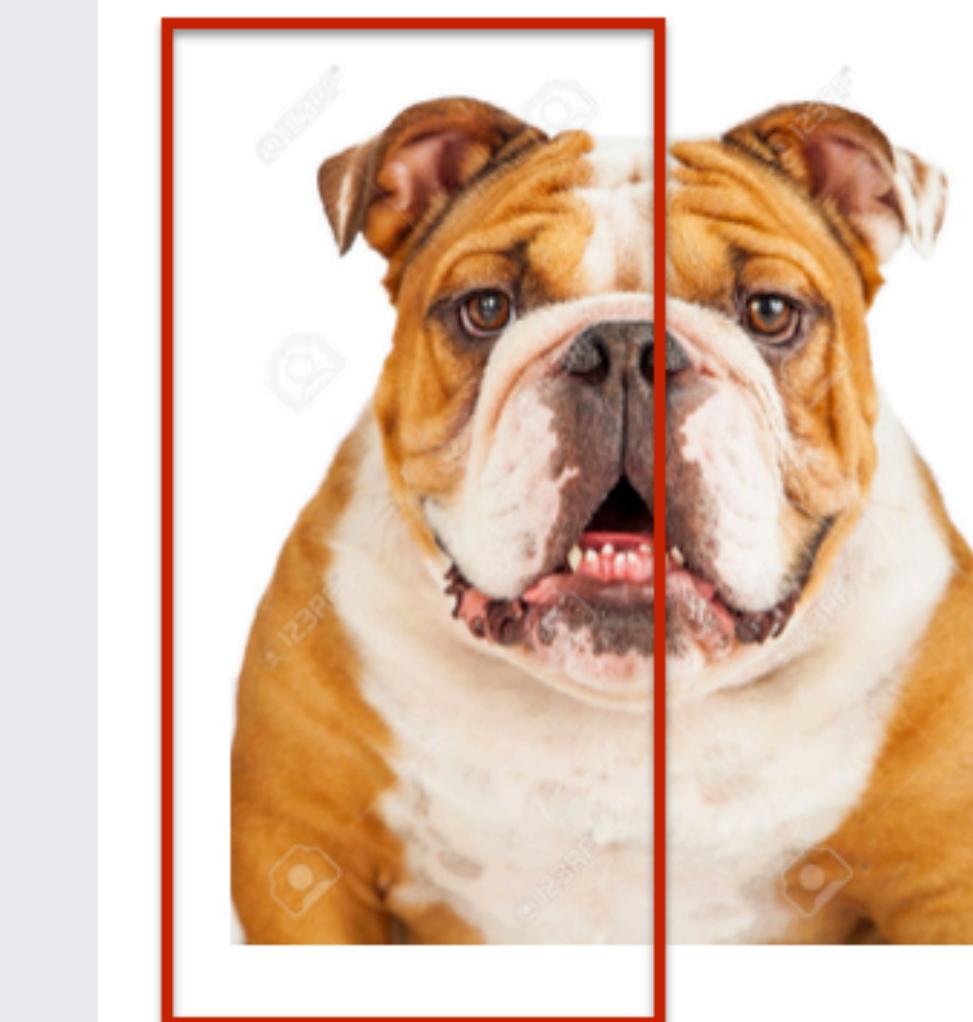
EN MAGDALENA EN 2018

VIENES?

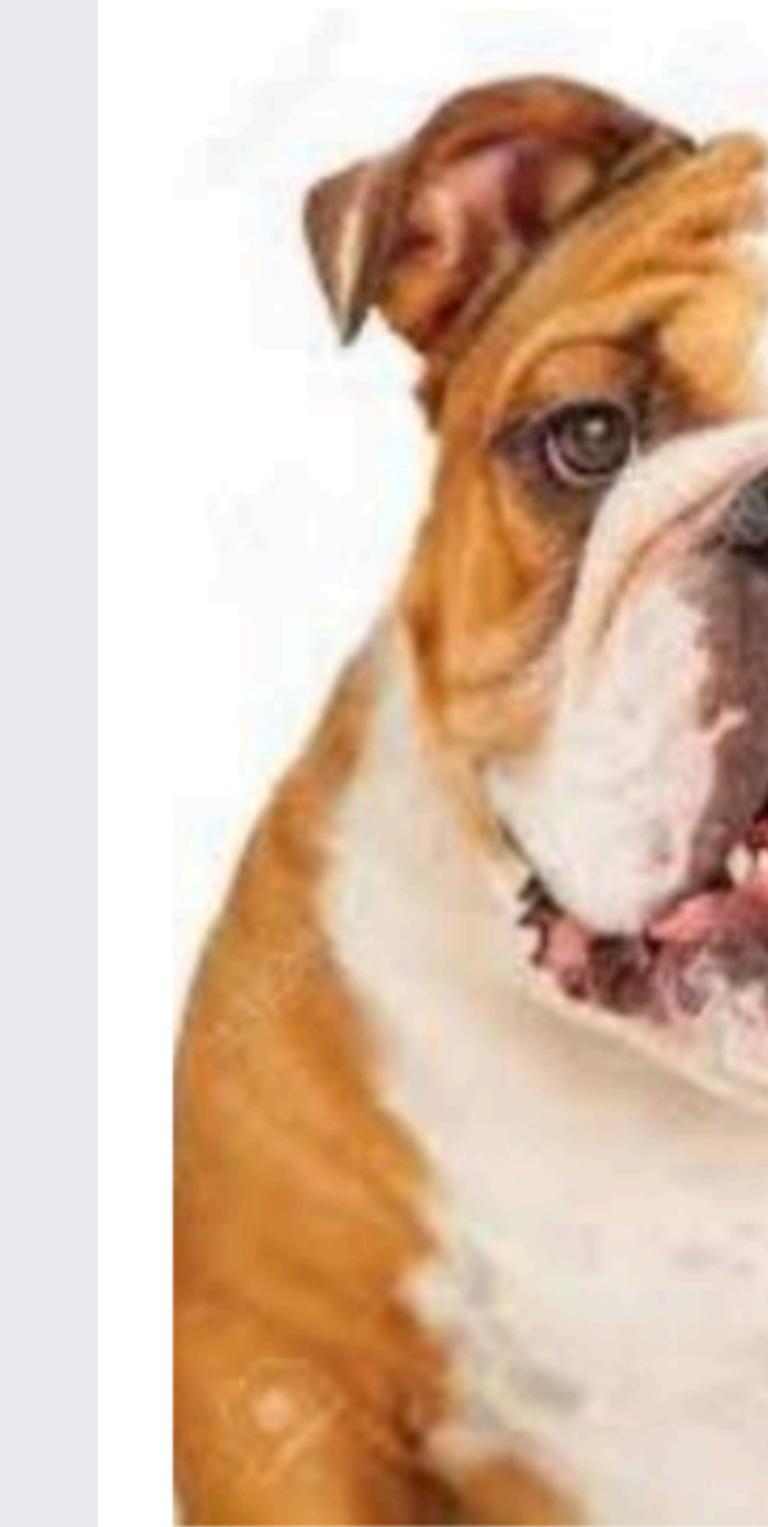
VIENES?

Challenges

- Sizes, fonts, orientations
- Languages
- Scale
- Efficiency

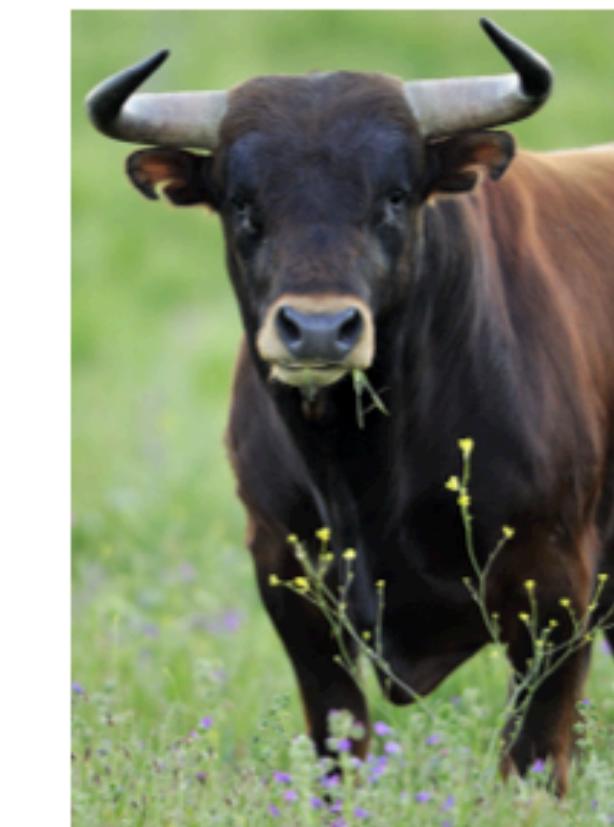


Classes for image

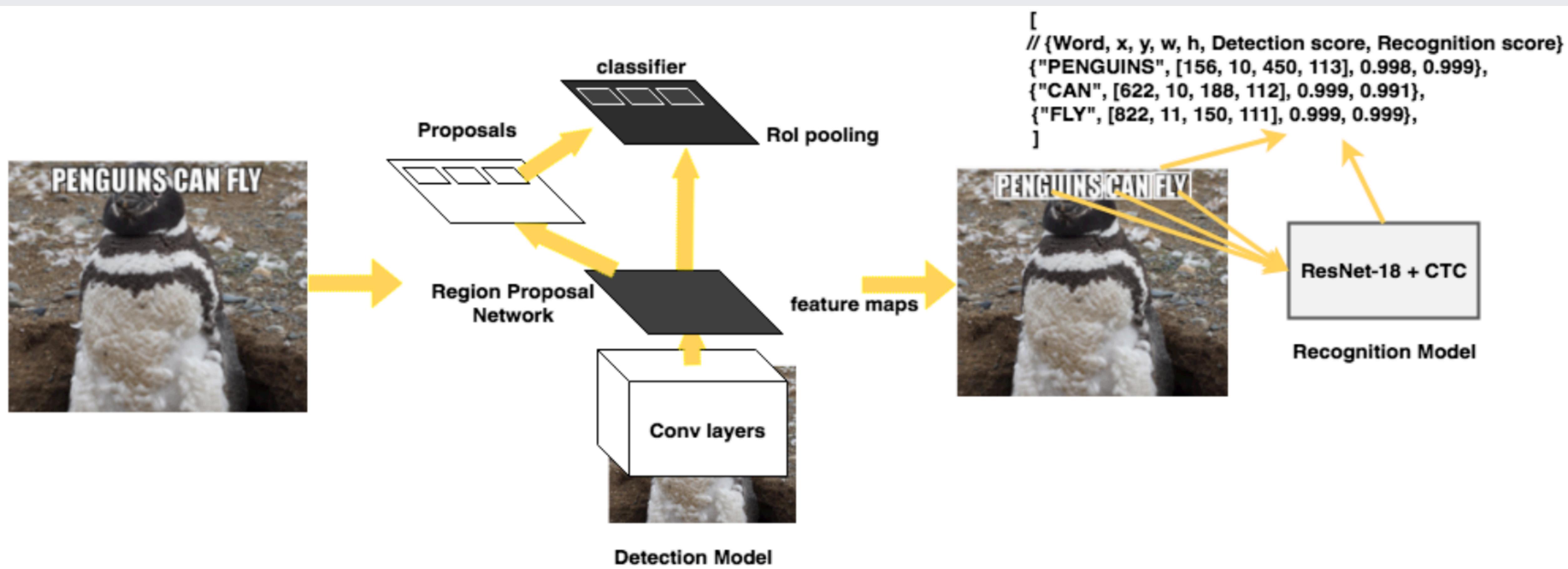


bulldog dog petshop pet english
puppy petfood bull dogfood creche
microchip haldol doggie peludo kennel

bulldog

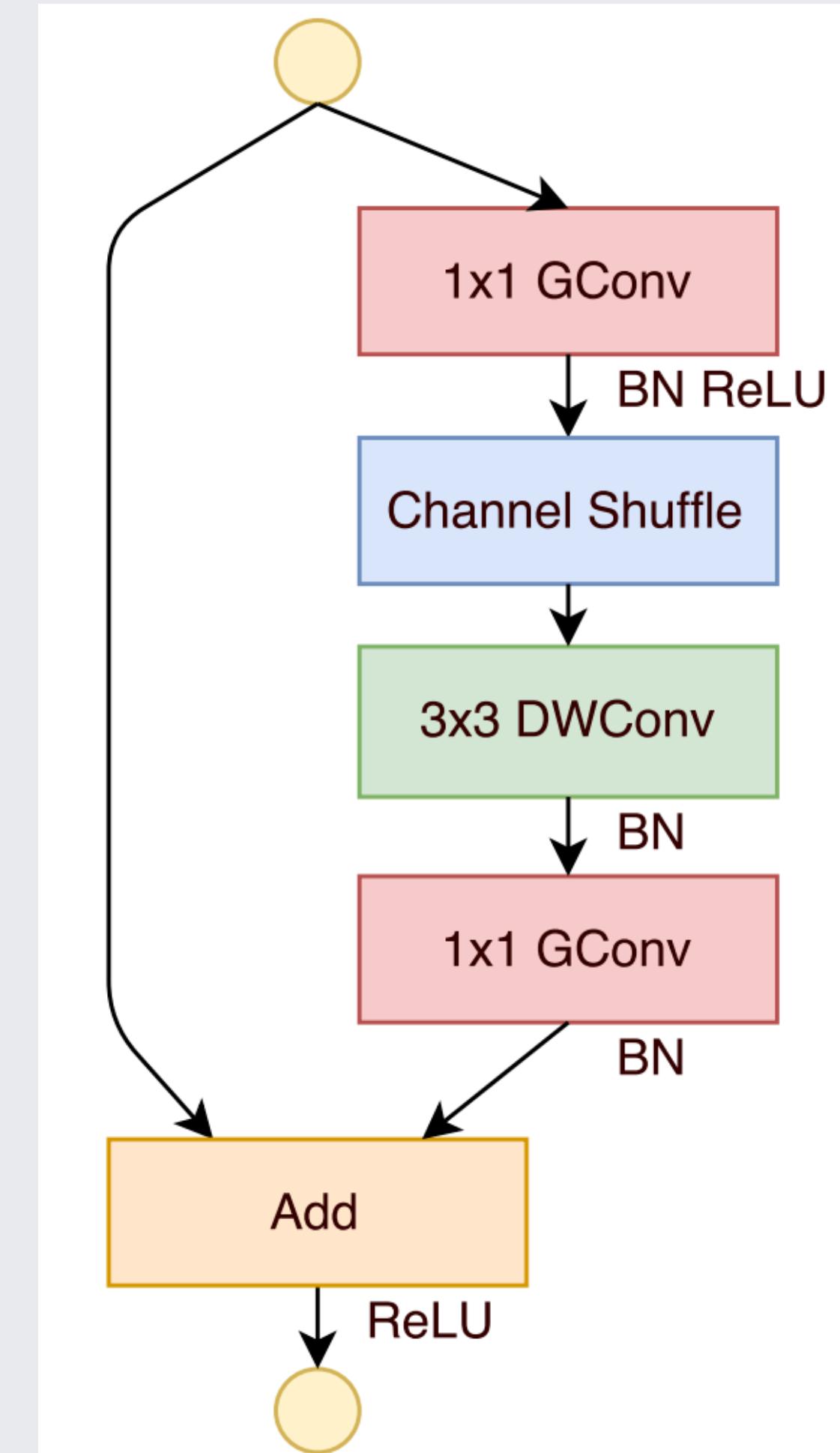
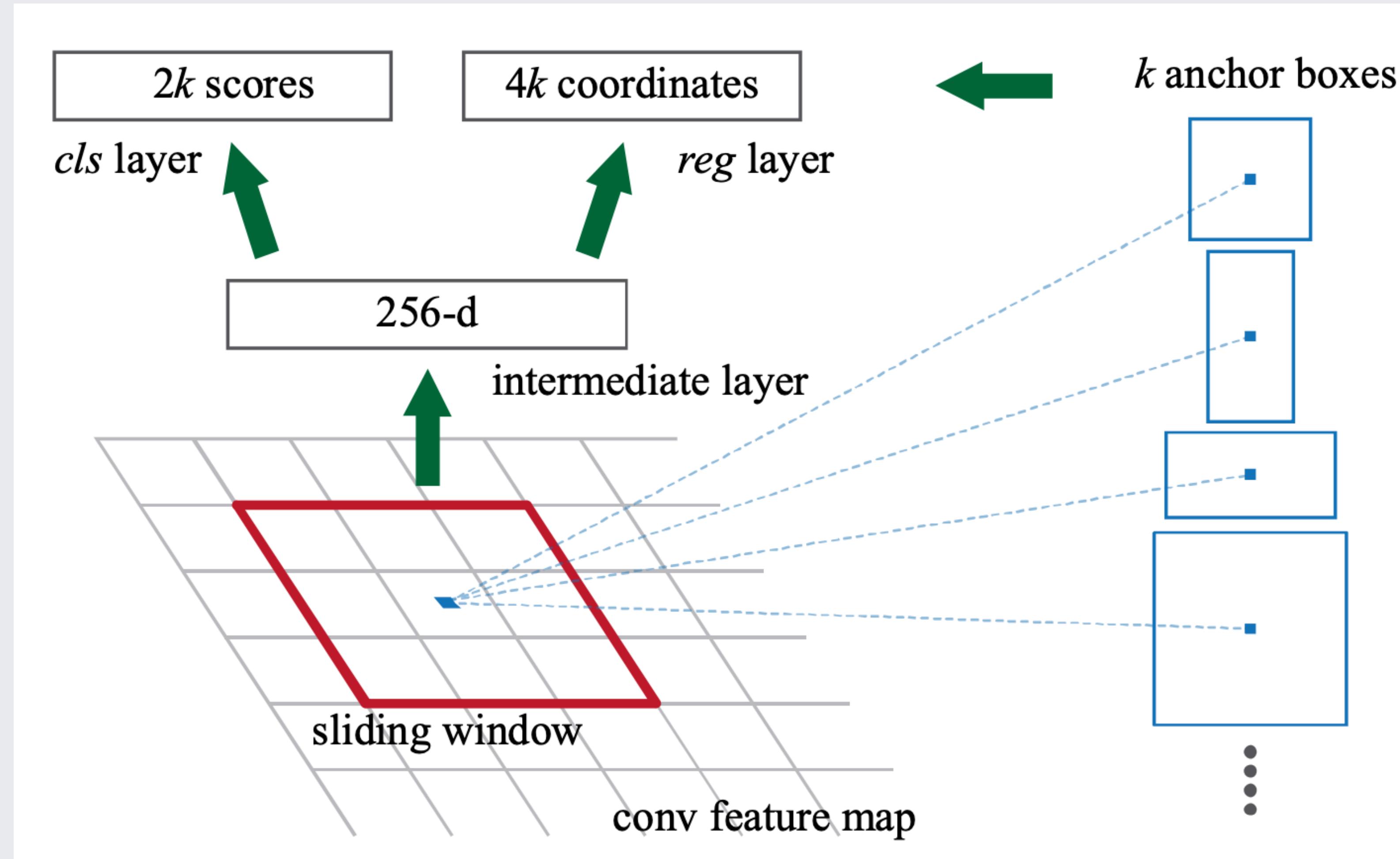


Architecture



Large scale system for text detection and recognition in images, KDD 2018,
Viswanath Sivakumar, Albert Gordo, Fedor Borisyuk

Text Detection



Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, Ren et al.

ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices, Zhang et al.

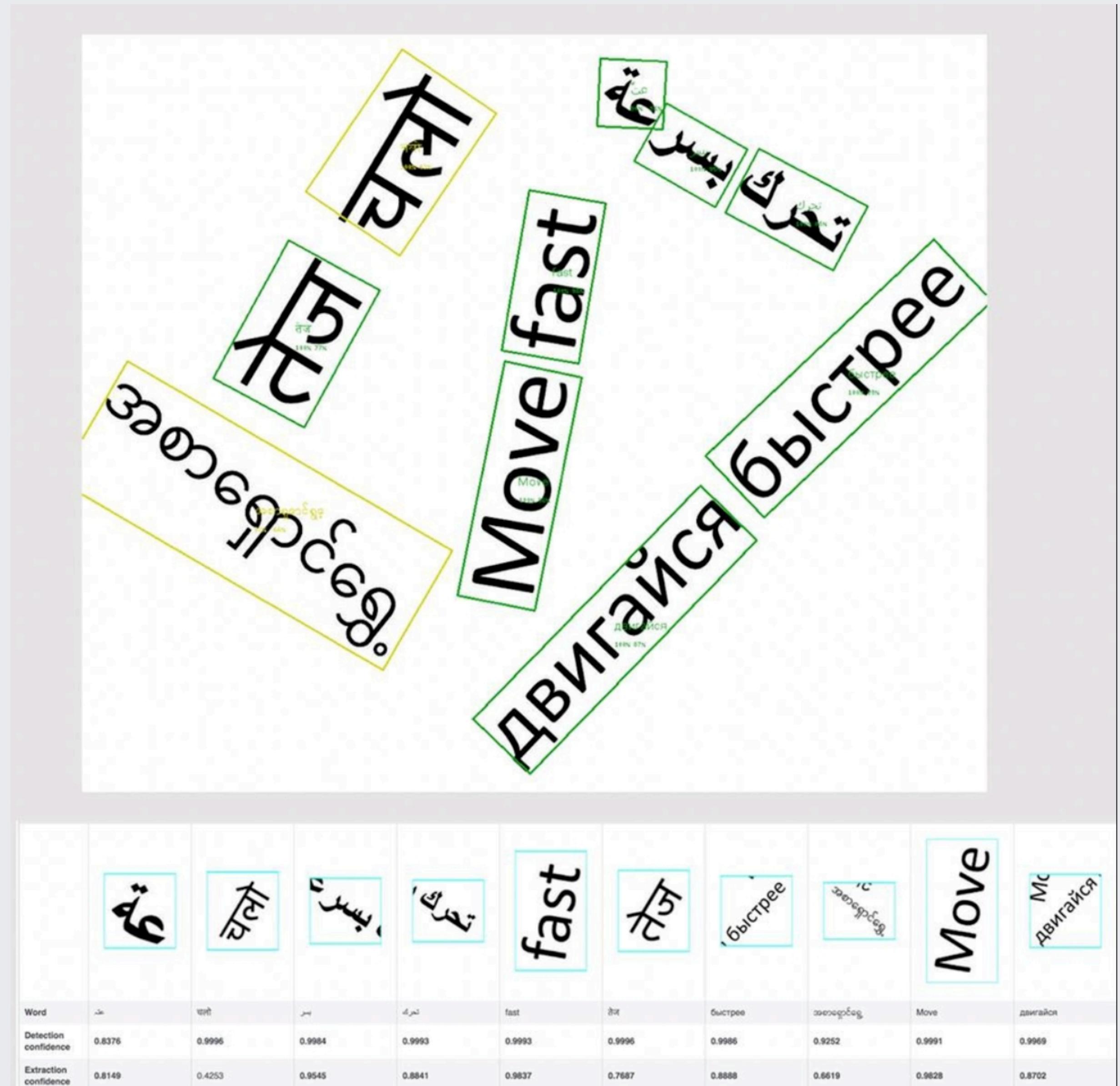
Orientations

The diagram illustrates the detection and orientation analysis of text in an image. The text 'Move fast' and its Russian translation 'двигайся быстрее' are shown in various orientations. Colored boxes (red, green, yellow) highlight specific words or parts of words, indicating their detected orientation and confidence.

Below the diagram is a table summarizing the word-level detection and extraction confidence scores:

Word	کی	لے	لے	لے	لے	fast	سے	Move	بے	میں	دھیاں
Detection confidence	0.7507	0.8731	0.9578	0.8830	0.9578	0.9006	0.9402	0.9900	0.9842		
Extraction confidence	0.7739	0.7489	0.6879	0.8583	0.7650	0.2286	0.4658	0.4614	0.2924		

Orientations



Improving Rotated Text Detection with
Rotation Region Proposal Networks,
Huang et al.

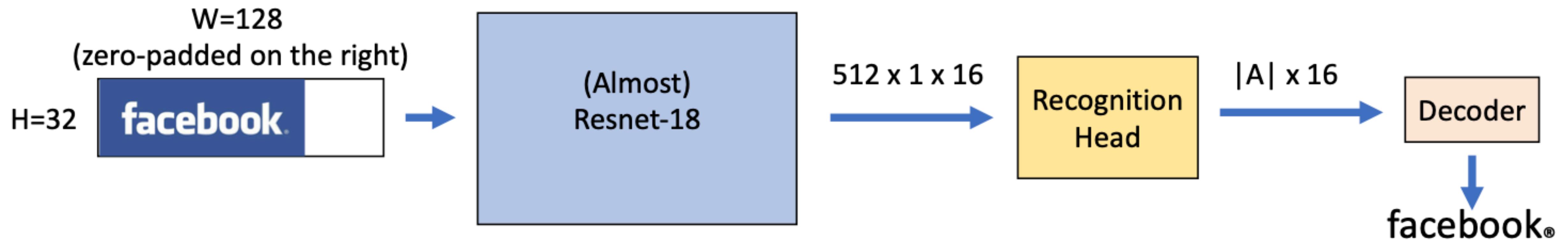
Text Recognition



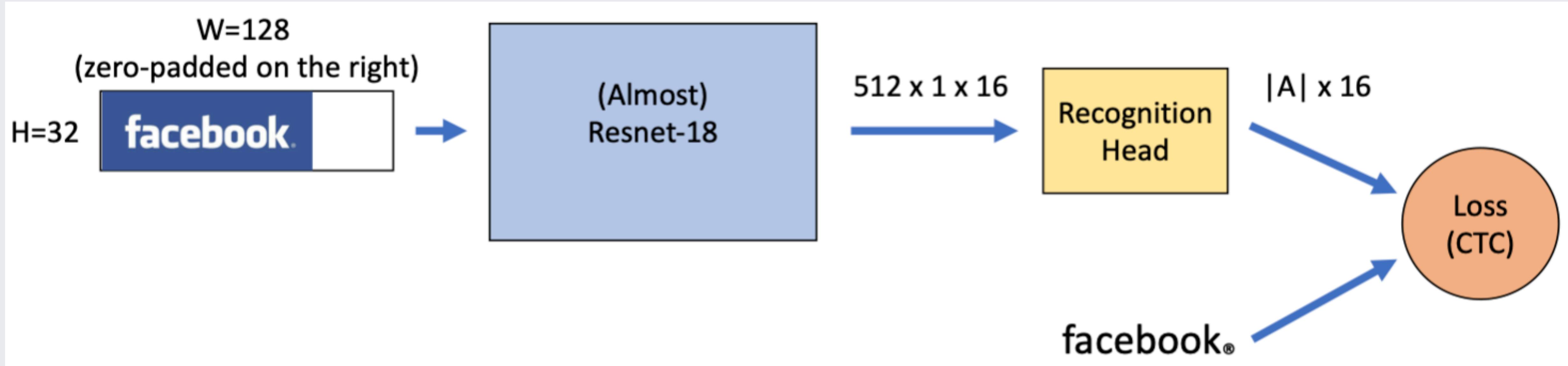
Text Recognition



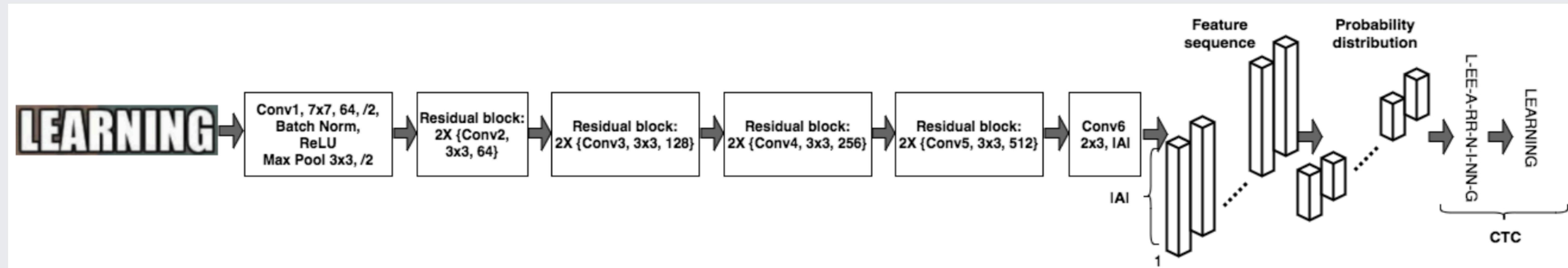
Text Recognition



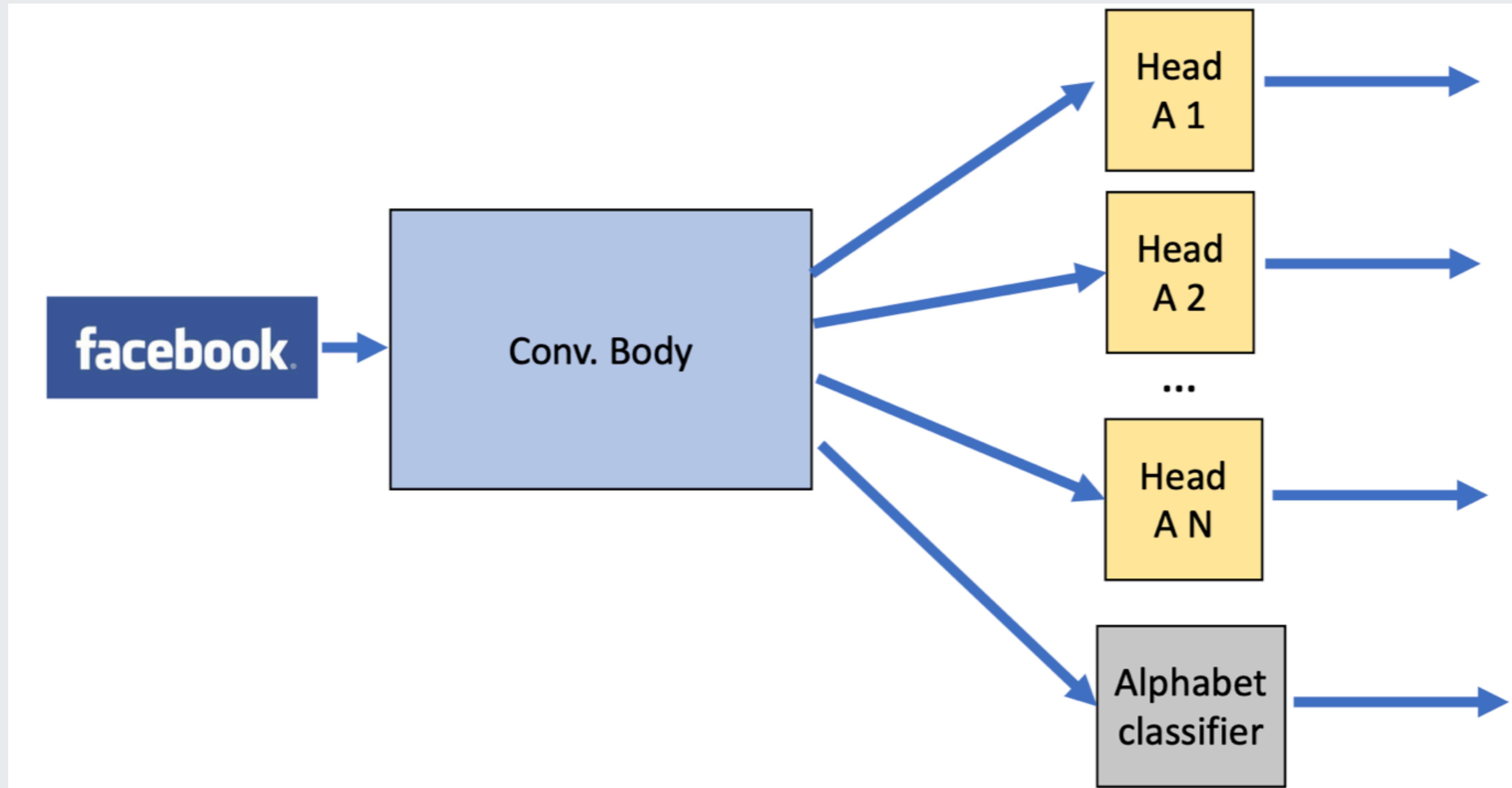
Text Recognition



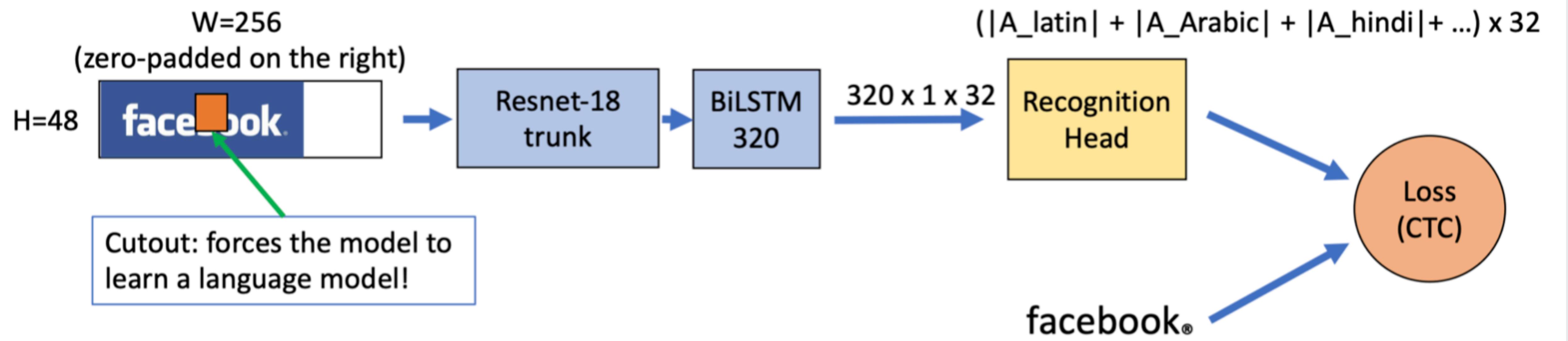
Text Recognition



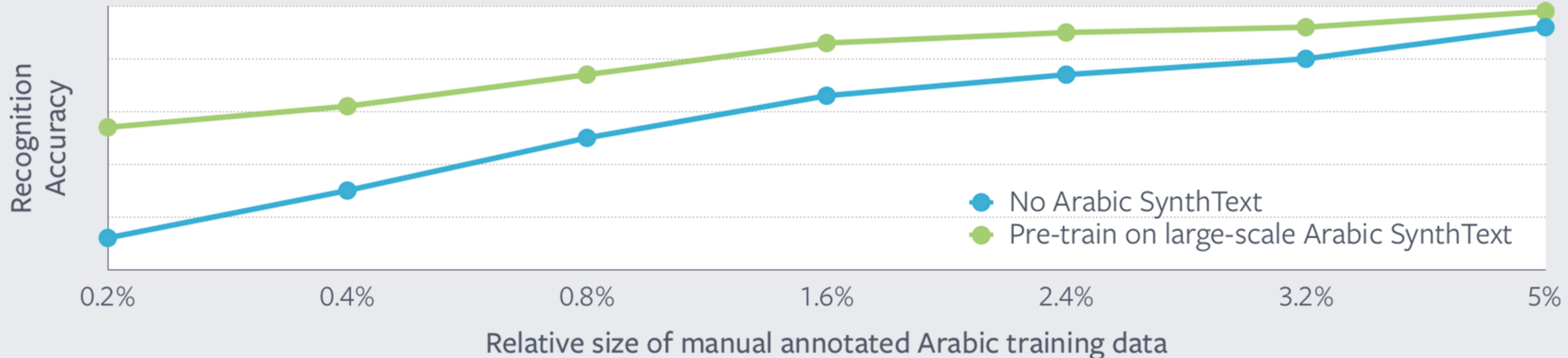
Multilingual



Multilingual



Synthetic Data



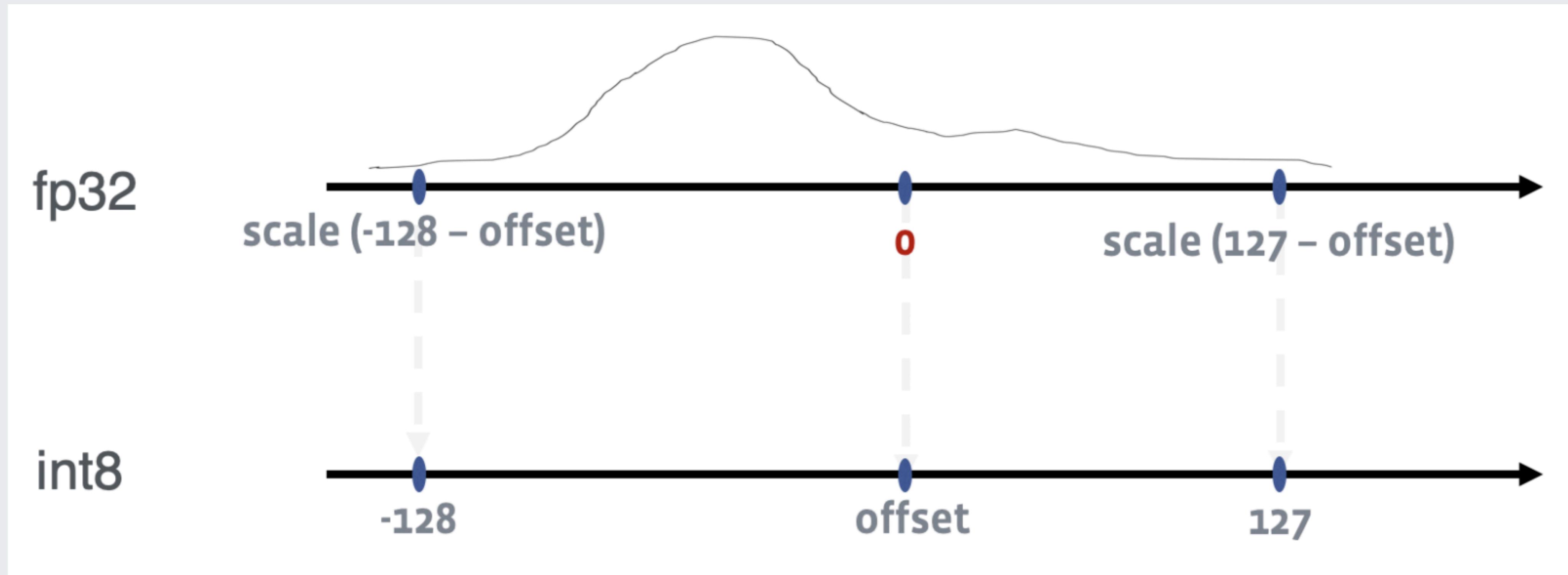
Inference

1B images/day x 5 sec/image = Lots of servers!

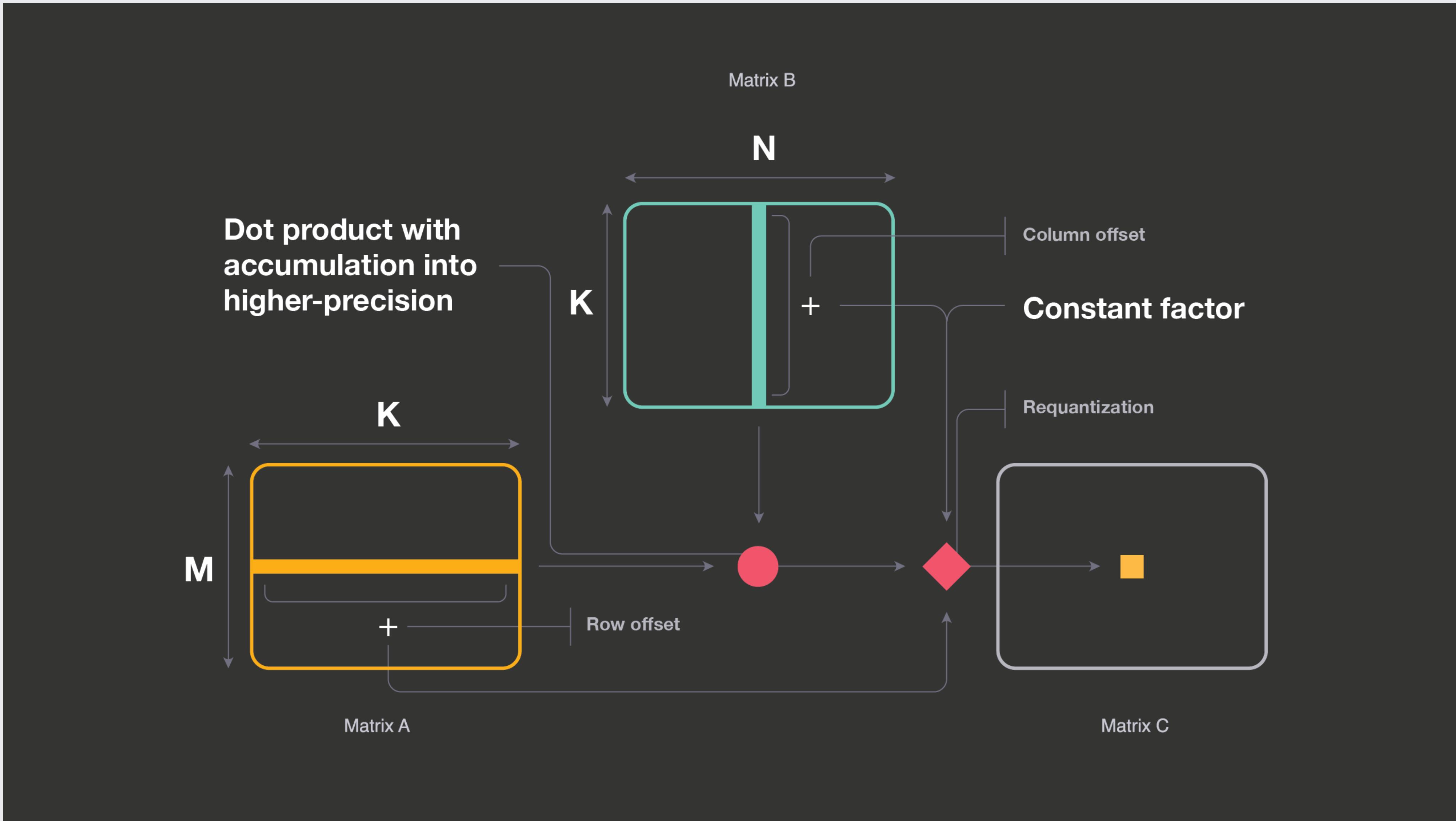
Int8 Quantization

Quantization: $x_q = \text{clip}(\text{round}(x/\text{scale}) + \text{offset}, -128, 127)$

De-quantization: $x = \text{scale} \times (x_q - \text{offset})$



Int8 Quantization



But what about accuracy?

- No quantization error for 0
- Fuse Convolution and ReLU
- L2 Error Minimization vs Min-Max
- Don't quantize the first layer

Initial accuracy gap: 5%

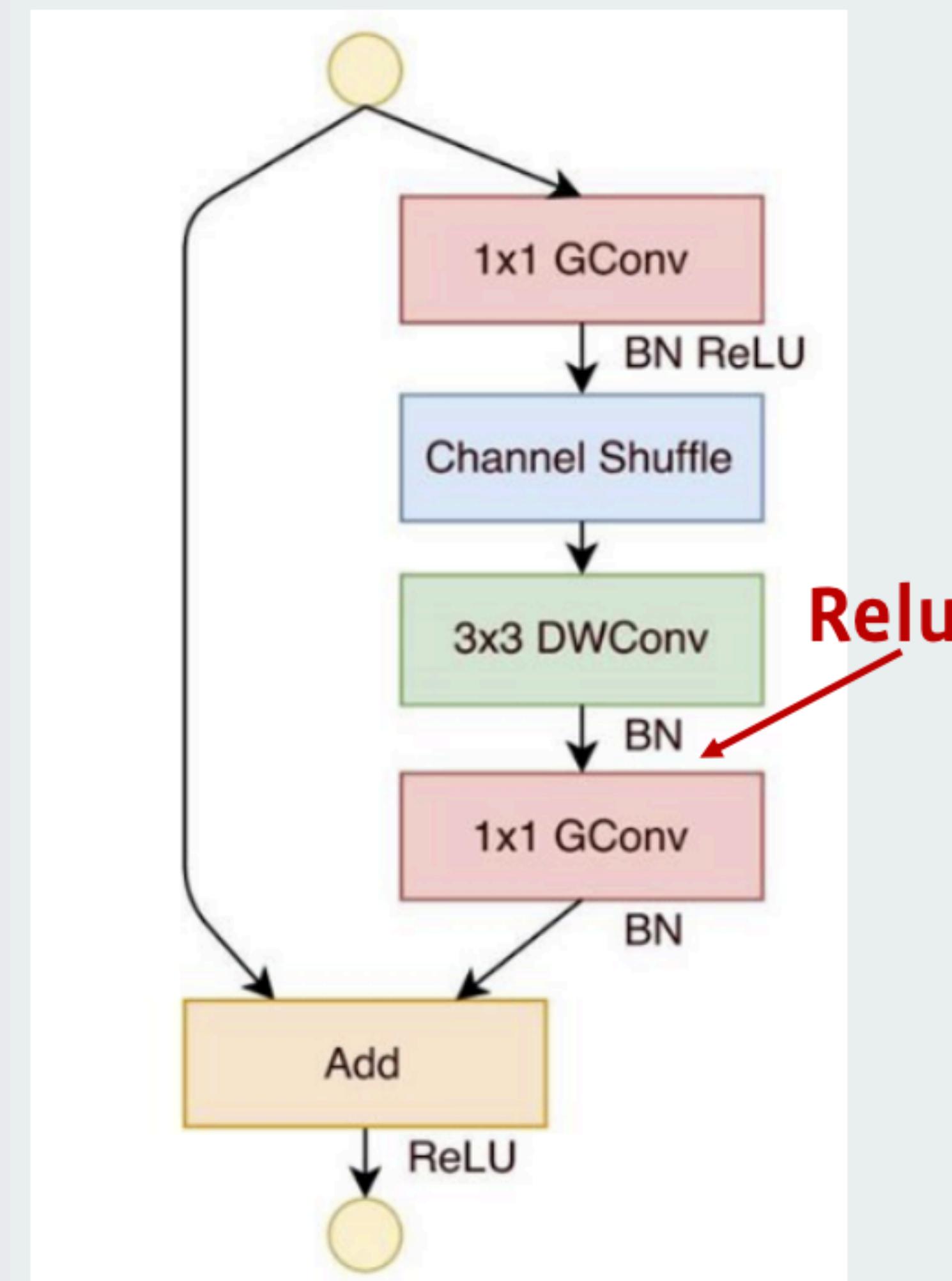
After: 0.2%

Model Co-Design

Outlier-aware quantization

- Int8 Quantization with 16-bit accumulation
- Further CPU speedup

Co-design your model and efficiency optimizations together!



ShuffleNet

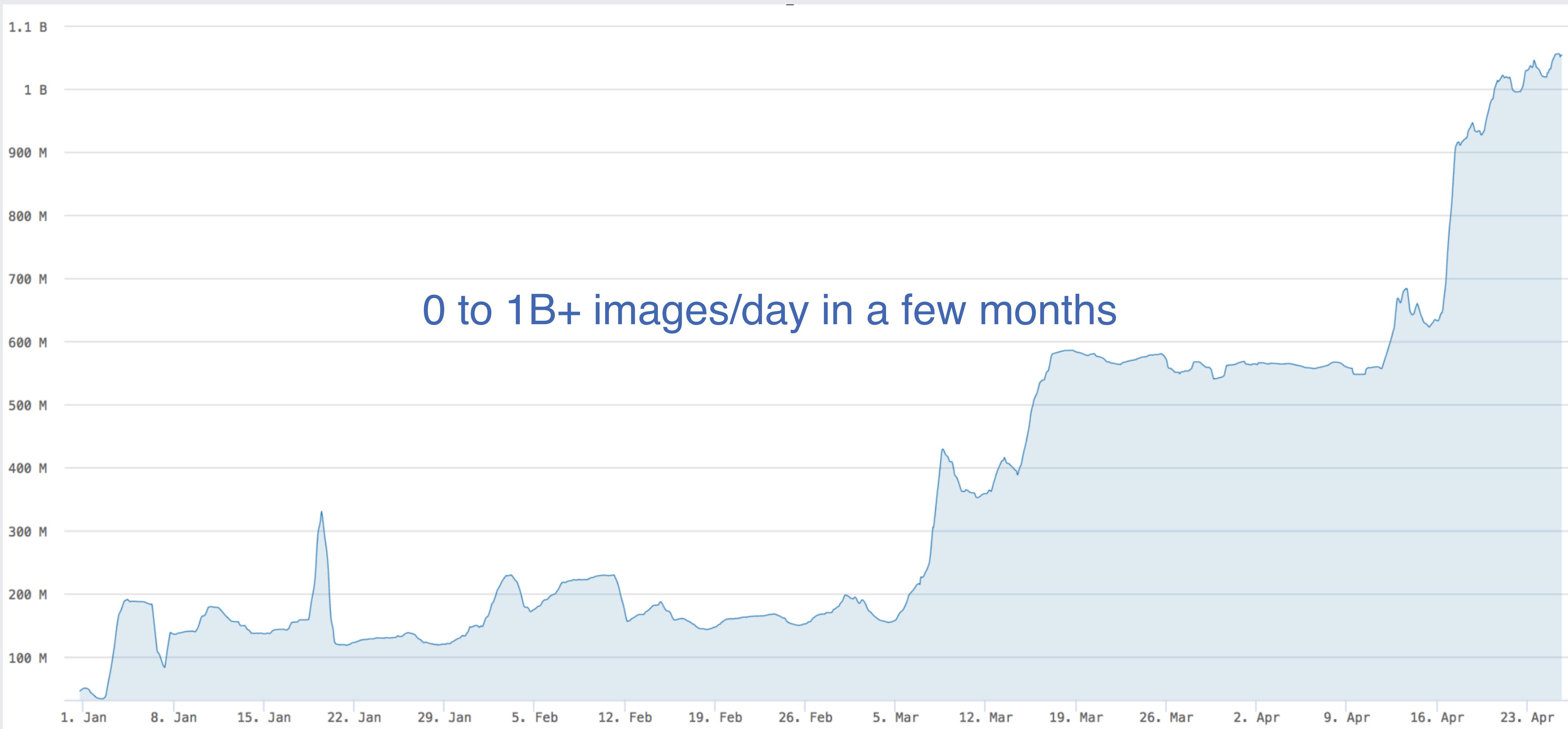
Int8 Quantization



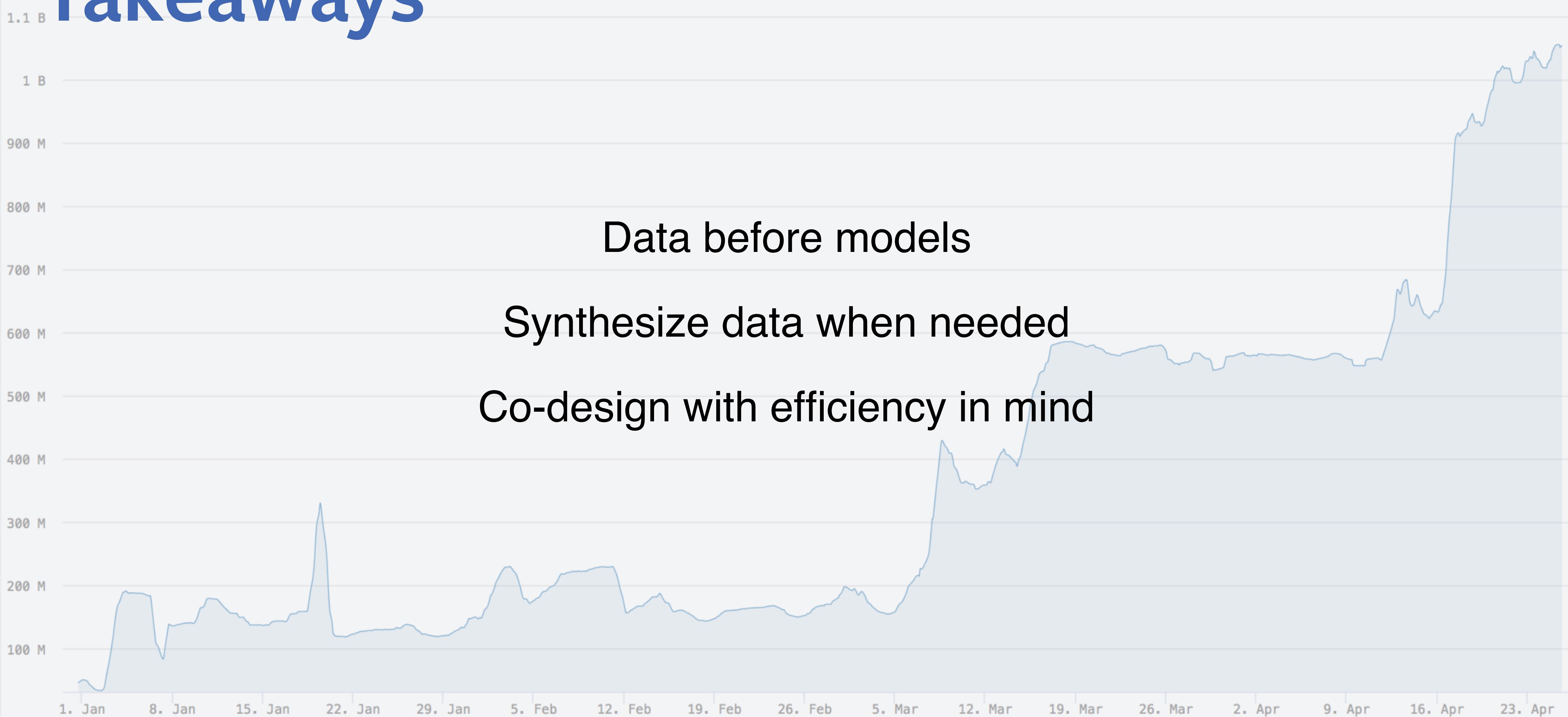
2.4x images/server

2x faster

github.com/pytorch/FBGEMM



Takeaways



Takeaways

Data before models

Synthesize data when needed

Co-design with efficiency in mind

1. Jan 8. Jan 15. Jan 22. Jan 29. Jan 5. Feb 12. Feb 19. Feb 26. Feb 5. Mar 12. Mar 19. Mar 26. Mar 2. Apr 9. Apr 16. Apr 23. Apr

Thank You!