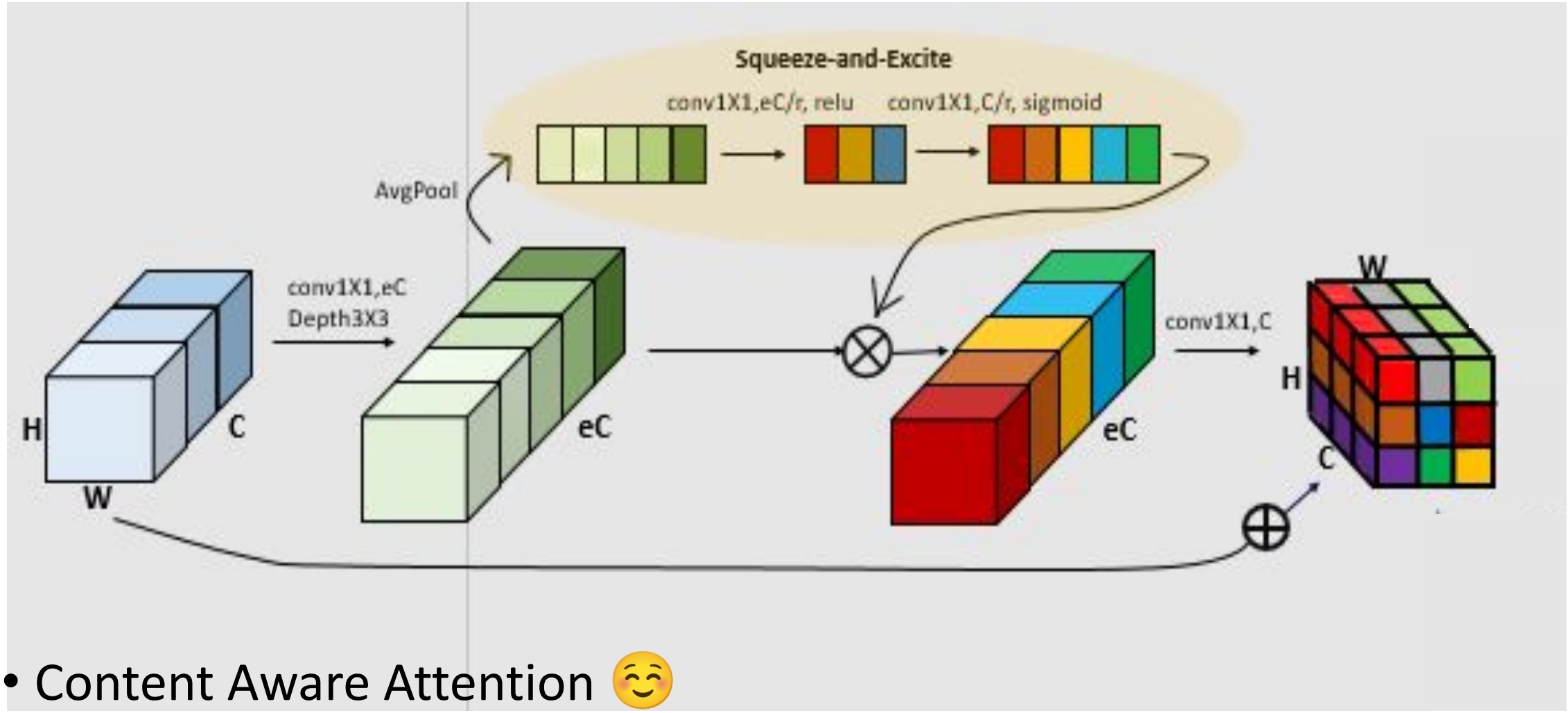
A SIMPLE APPROACH TO **IMAGE TILT CORRECTION** USING **SELF-ATTENTION MOBILENETS** FOR SMARTPHONES

- SELF-ATTENTION MOBILENET
- TRAINING PIPELINE FOR IMAGE TILT CORRECTION

SELF-ATTENTION MOBILENET

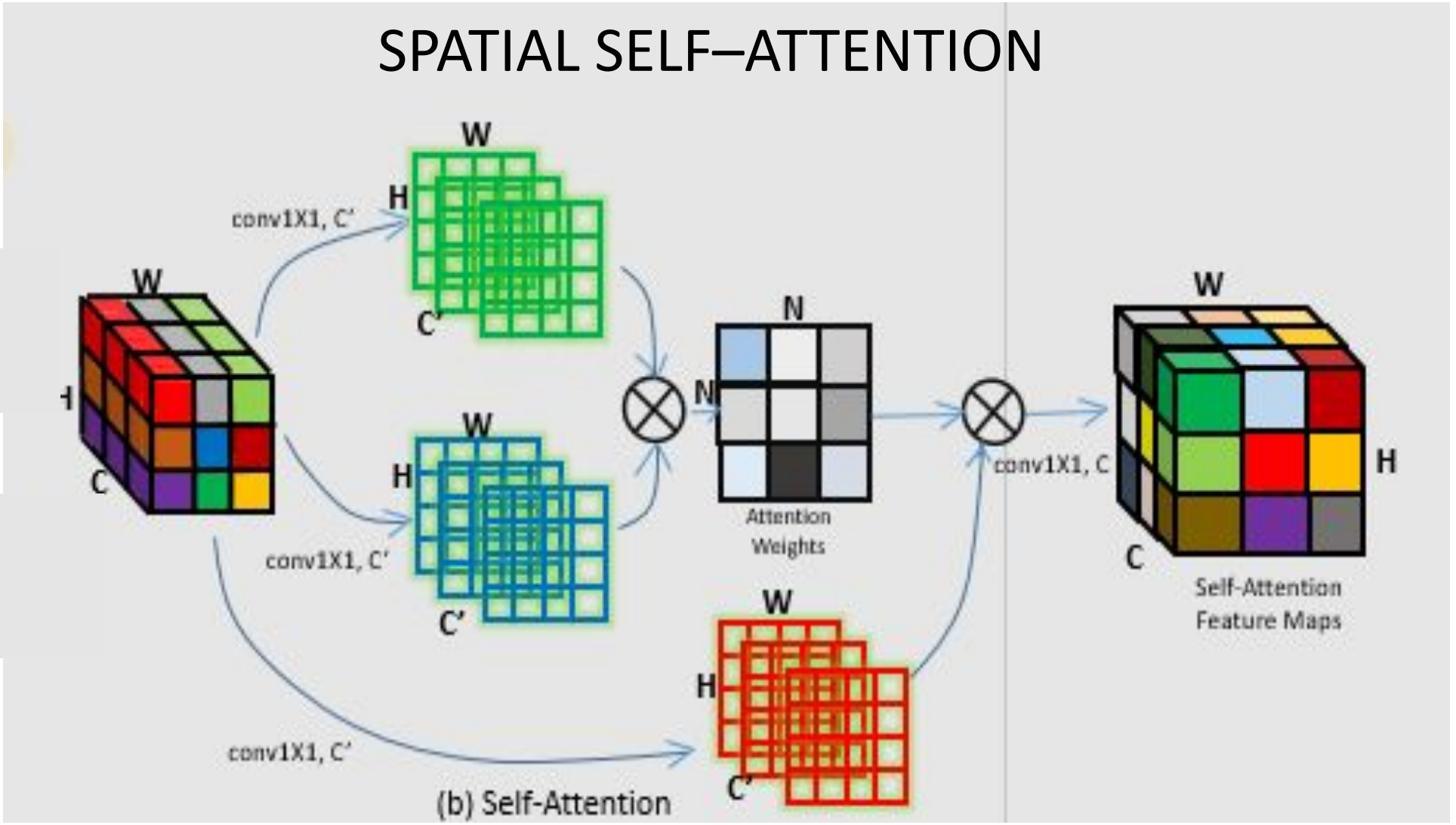
- SQUEEZE-AND-EXCITE/CHANNELWISE ATTENTION (MobileNetV3)
- SPATIAL SELF-ATTENTION (PROPOSED)

SQUEEZE-AND-EXCITE or CHANNELWISE SELF–ATTENTION

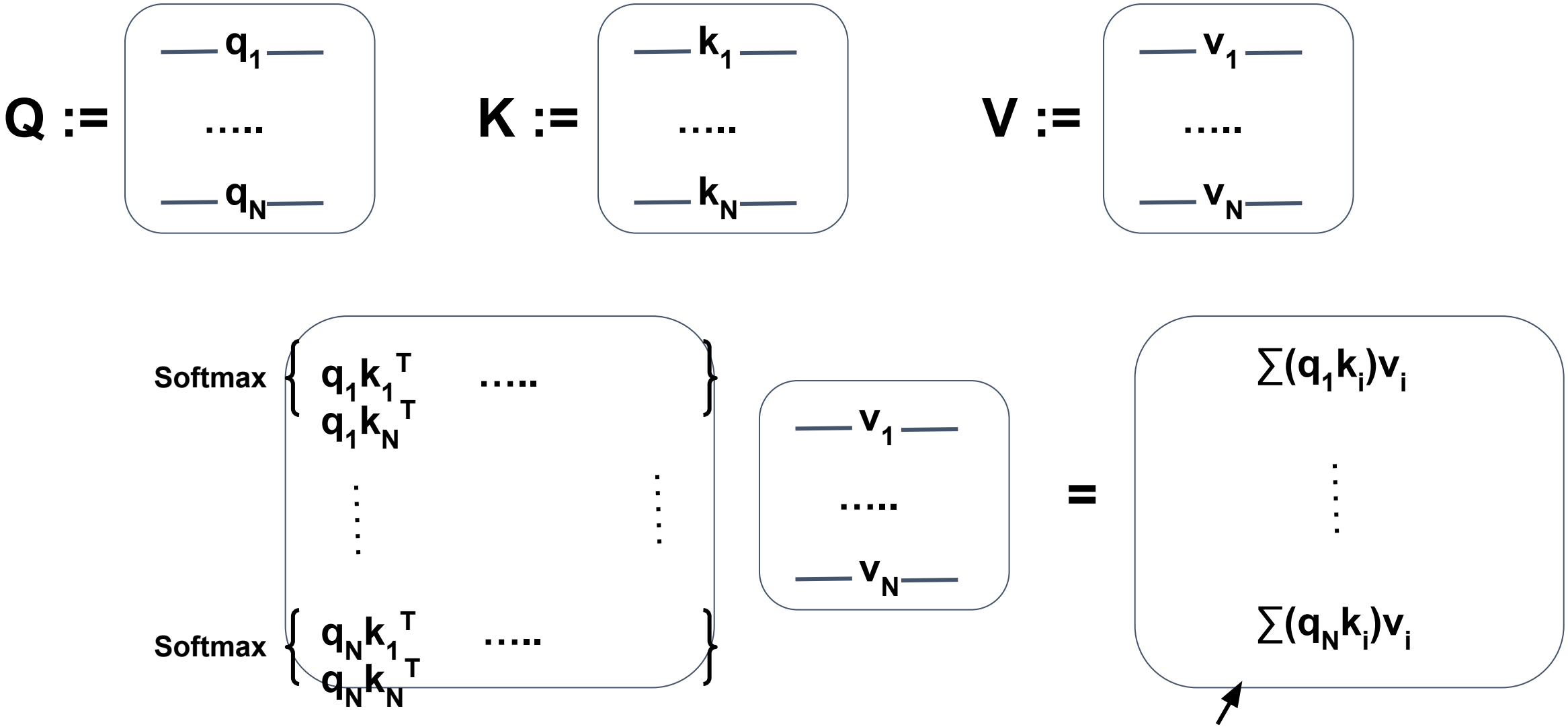


- Content Aware Attention ☺️
- Global Average Pooling ☹️
- No explicit far-away context ☹️

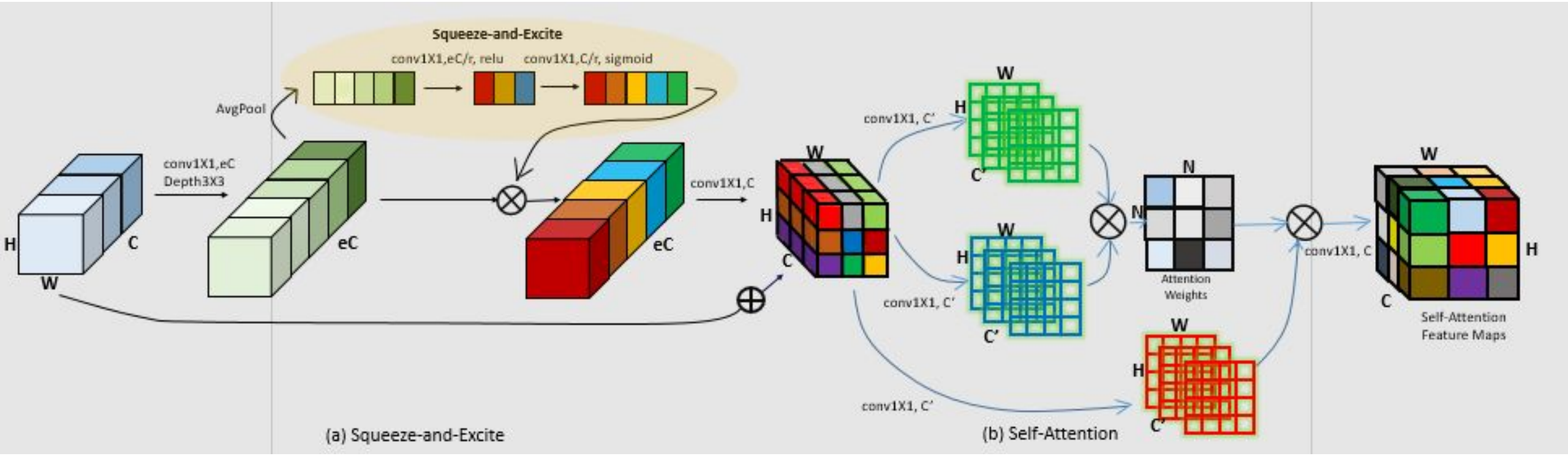
SPATIAL SELF–ATTENTION



- Long-range contextual information ☺
- Relative ordering between image features/pixels ☺


Q :=
q_1
.....
q_N
K :=
k_1
.....
k_N
V :=
v_1
.....
v_N
Softmax
$q_1 k_1^T$ ..... $q_1 k_N^T$
$q_N k_1^T$ ..... $q_N k_N^T$
Softmax
v_1
.....
v_N
=
$\sum(q_1 k_i) v_i$
$\sum(q_N k_i) v_i$
Sum of weighted values w.r.t how q_N attends to other regions.

SELF–ATTENTION MOBILENET
SQUEEZE-AND-EXCITE + SPATIAL SELF-ATTENTION



- Content Aware Attention ☺️
- Long-range contextual information ☺️
- Computationally Efficient ☺️

IMAGE TILT CORRECTION

INSIGHT-1

True “zero” is ambiguous (quite often)

-1°

0°

1°

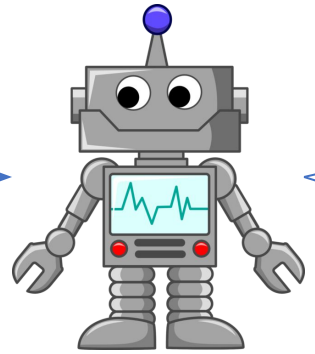
INSIGHT-2

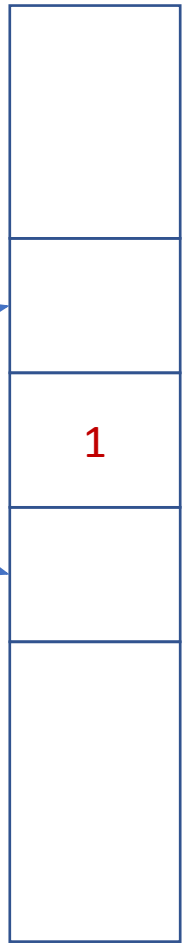
Scarcity of Labelled Data

TILTED IMAGE

ANNOTATED UPRIGHT IMAGE

HOW TO TRAIN?





-1

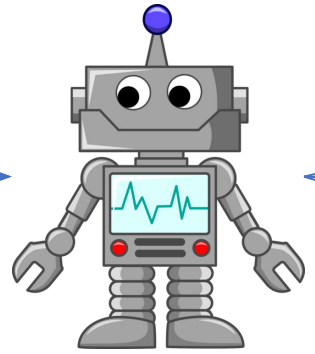
0

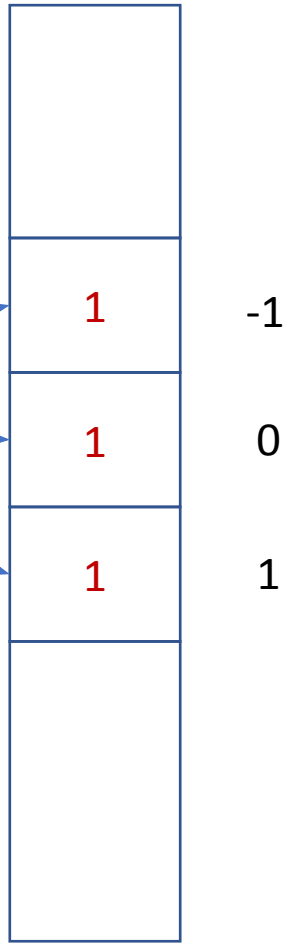
1

Single Label Classification

- Good Training Performance for deep learning tasks in general
- Too Strict in our case

HOW TO TRAIN?



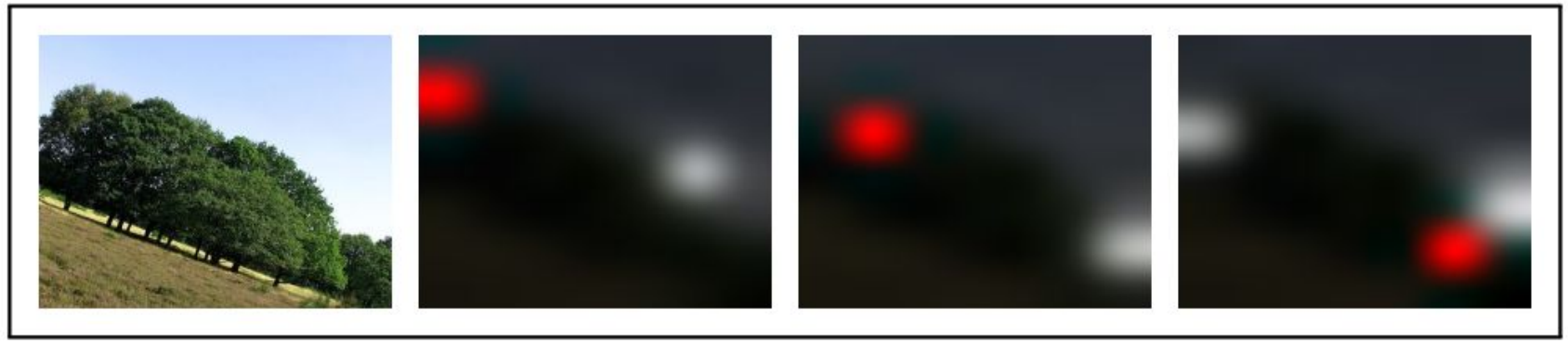


Multi Label Classification

- Good Training Performance
- Handles Upright/True-0 Image ambiguity
- Low angle error in prediction

1 1.*a* 1.*b* 1.*c*

2 2.*a* 2.*b* 2.*c*



3 3.*a* 3.*b* 3.*c*

~25M

~4.5M

Model	NYU-V1		ADE20K		SUN397	
	Acc(%) ↑	AE° ↓	Acc(%) ↑	AE° ↓	Acc(%) ↑	AE° ↓
MobileNetV3	88.02	15.79	87.68	16.84	85.97	5.06
ResNet-50	94.59	4.67	**97.84**	**3.09**	**93.67**	**3.98**
SA-MobileNet	**98.53**	**3.45**	96.77	3.45	92.39	4.27

Table 2: Evaluation accuracies and angle errors of the MobileNetV3, ResNet-50, and SA-MobileNet models on various datasets with the proposed tilt angle detection approach. Acc: Accuracy (%) and AE: Angle Errors(°). ↑ / ↓ indicates that higher/lower is better respectively.

Model	Latency(↓) (milliseconds)	Parameters(↓) (millions)
MobileNetV3	79	**4.2**
SA-MobileNet	**75**	4.5

Table 3: Tflite models were tested on Snapdragon 750, Octa core (2x 2.2 GHz, 6x 1.8 GHz) for latency measurements.

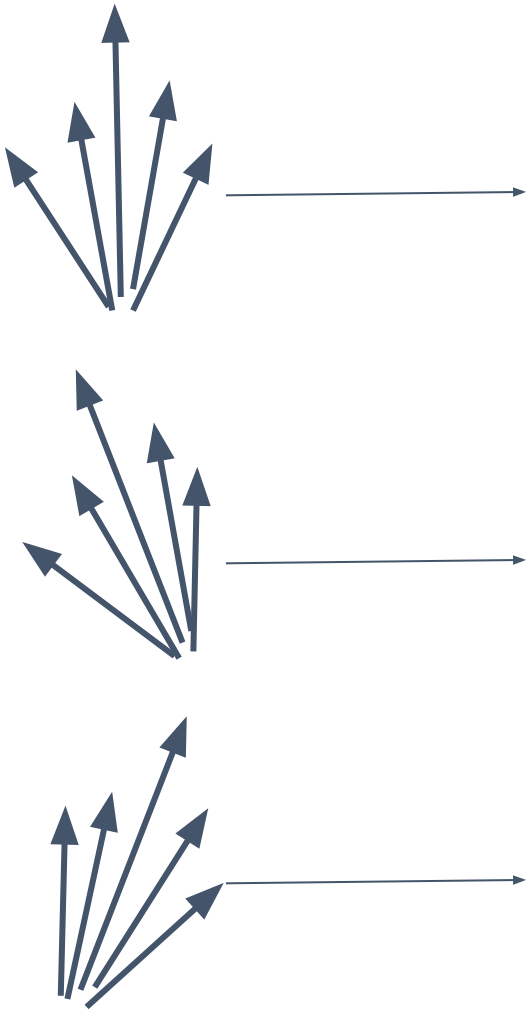
Model	Angle Error° (↓)
MobileNetV3	21.07
SA-MobileNet	**15.53**

Table 4: Regresion loss on ADE20k dataset trained with angle loss function Eq.8.

Previous Works	**Accuracy** (%) ↑	**Angle Error** (°) ↓
Ciocca *et al.* [3] (LPB-based featutes)	71.87	-
CNN + Fuzzy Edge Detection	85.21	-
Fischer *et al.* [6] (AlexNet)	-	21.23
Maji *et al.* [13] (Xception)	-	7.89
MobileNetV3 (baseline)	85.97	5.06
ResNet-50 (baseline)	**93.67**	**3.98**
SA-MobileNet (proposed)	92.39	4.27

Table 1: Accuracies and angles errors of various baseline methods on SUN397 dataset. ↑ / ↓ indicates that higher/lower is better respectively.

Training Phase



-2, -1, 0, 1, 2

-4, -3, -2, -1, 0

0, 1, 2, 3, 4

Testing Phase

Tilt Correction in both the cases (Empirically observed)

THANK YOU  !!