Supplement to Gamifying Math Education using Object Detection

Yueqiu Sun
Tangible Play
Palo Alto, CA 94306
yueqiu@tangibleplay.com

Rohitkrishna Nambiar
Tangible Play
Palo Alto, CA 94306
rohit@tangibleplay.com

Vivek Vidyasagaran
Tangible Play
Palo Alto, CA 94306
vivek@tangibleplay.com

A Appendix

A.1 Network Architecture

Table 1 shows the network configuration for models with Mobilenet backbone and VGG backbone.

Note that "Conv dw" refers to depthwise convolution, and "Conv pw" refers to pointwise convolution. $output32$ and $output16$ refers to the outputs produced by the backbone with regard to image resolution 32 x 32 and 16 x 16 respectively.

Table 1: Mobilenet Backbone and VGG Backbone Architectures

Layers	Output Size	Mobilenet Backbone	VGG Backbone
Conv Block	128 x 128	3 x 3 conv	
Conv Block	128 x 128	[3 x 3 Conv dw 1 x 1 Conv pw] x 2	3 x 3 conv
Pooling Block	64 x 64	[1 x 1 Conv 3 × 3 max pool, stride 2]	3 × 3 max pool, stride 2
Conv Block	64 x 64	[3 x 3 Conv dw 1 x 1 Conv pw] x 3	3 x 3 conv
Pooling Block($output32$)	32 x 32	[1 x 1 Conv 3 × 3 max pool, stride 2]	3 × 3 max pool, stride 2
Conv Block	32 x 32	[3 x 3 Conv dw 1 x 1 Conv pw] x 3	3 x 3 conv
Pooling Block($output16$)	16 x 16	[1 x 1 Conv 3 × 3 max pool, stride 2]	3 × 3 max pool, stride 2

Outputs from the backbone, i.e.,$output32$ and $output16$ are concatenated and forwarded to the classification head, the orientation head and the regression head respectively. The network configuration for each head is shown in Table 2.

35th Conference on Neural Information Processing Systems (NeurIPS 2021) Workshop on Math AI for Education (MATHAI4ED).

Note that $nbb32$ and $nbb16$ refer to the number of default bounding box with regard to image resolution 32 x 32 and 16 x 16 respectively. $ncls$ refers to the number of classes(types of shapes). $nori$ refers to the number of bins for orientations.

Table 2: Configurations for the classification head, the orientation head and the regression head

	Classification head	Orientation head	Regression head
Layers		3 x 3 conv + reshape	
Output Size	$(nbb32 + nbb16) \times ncls$	$(nbb32 + nbb16) \times nori$	$(nbb32 + nbb16) \times 4$

A.2 Shape manipulatives

Figure 1: Tangible shape manipulatives

In the game, the following geometric pieces in different quantities are used to form food ingredients. We have a green quarter circle, a pink semicircle, teal square, lime rectangle, orange isosceles triangle, red scalene right triangle, green obtuse isosceles triangle, teal equilateral triangle, dark blue rhombus and purple trapezoid.