Appendix

A Neural Path Framework For CNN

Indexing: The weights of layers $l \in [d_{\text{cv}}]$ are denoted by $\Theta(i_{\text{cv}}, i_{\text{in}}, i_{\text{out}}, l)$ and for layers $l \in [d_{\text{fc}}]+d_{\text{cv}}$ are denoted by $\Theta(i_{\text{in}}, i_{\text{out}}, l)$. The pre-activations, gating and hidden unit outputs are denoted by $q_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, l)$, $G_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, l)$, and $z_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, l)$ for layers $l = 1, \ldots, d_{\text{cv}}$. i_{in} and i_{out} are used to index the input and the output filters. i_{fout} is used to denote the index of hidden unit (in the feature dimension) within the input and output filters.

Shapes: Appendix A shows the shapes of the tensors in the convolutional layers of a 1-dimensional circular CNN considered in this paper. Here, the input is a 1-dimensional tensor given by $x \in \mathbb{R}^{d_{\text{in}}}$. The hidden nodes in a given convolutional layer have a 2-dimensional shape of $d_{\text{in}} \times w$, where w is the number of filters in the layer. The weights of a given convolutional layer have 3-dimensional shape of $w_{\text{cv}} \times w \times w$, where $w \times w$ is because of the number of input filters times the number of output filters.

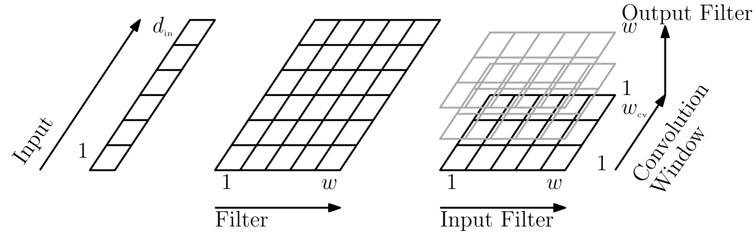


Figure 6: Shows the shape of the tensor.

A.0.1 Information Flow

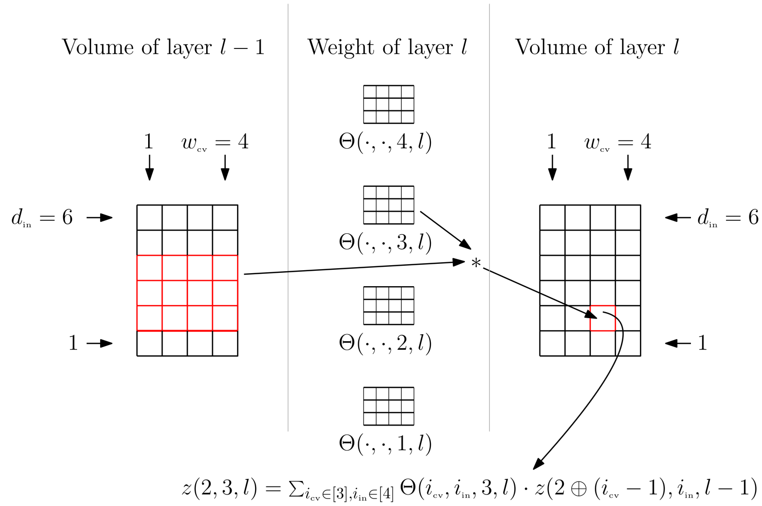
IL	:	$z_{x,\Theta}(\cdot, 1, 0)$	$=$	x
Convolutional Layers, $l \in [d_{\text{cv}}]$				
PA	:	$q_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, l)$	$=$	$\sum_{i_{\text{cv}}, i_{\text{in}}} \Theta(i_{\text{cv}}, i_{\text{in}}, i_{\text{out}}, l) \cdot z_{x,\Theta}(i_{\text{fout}} \oplus (i_{\text{cv}} - 1), i_{\text{in}}, l-1)$
GV	:	$G_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, l)$	$=$	$\mathbf{1}_{\{q_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, l)>0\}}$
HUO	:	$z_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, l)$	$=$	$q_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, l) \cdot G_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, l)$
GAP Layers, $l = d_{\text{cv}} + 1$				
HUO	:	$z_{x,\Theta}(i_{\text{out}}, d_{\text{cv}} + 1)$	$=$	$\sum_{i_{\text{fout}}} z_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, d_{\text{cv}}) \cdot G^{\text{pool}}_{x,\Theta}(i_{\text{fout}}, i_{\text{out}}, d_{\text{cv}} + 1)$
Fully Connected Layers, $l \in [d_{\text{fc}}] + (d_{\text{cv}} + 1)$				
PA	:	$q_{x,\Theta}(i_{\text{out}}, l)$	$=$	$\sum_{i_{\text{in}}} \Theta(i_{\text{in}}, i_{\text{out}}, l) \cdot z_{x,\Theta}(i_{\text{in}}, l-1)$
GV	:	$G_{x,\Theta}(i_{\text{out}}, l)$	$=$	$\mathbf{1}_{\{(q_{x,\Theta}(i_{\text{out}}, l))>0\}}$
HUO	:	$z_{x,\Theta}(i_{\text{out}}, l)$	$=$	$q_{x,\Theta}(i_{\text{out}}, l) \cdot G_{x,\Theta}(i_{\text{out}}, l)$
FO	:	$\hat{y}_\Theta(x)$	$=$	$\sum_{i_{\text{in}}} \Theta(i_{\text{in}}, i_{\text{out}}, d) \cdot z_{x,\Theta}(i_{\text{in}}, d-1)$

Table 2: Here IL, PA, GV, HUO, GL and FO are abbreviations for input layer, pre-activation, gating values, hidden unit output, GAP-layer and final output respectively.

B Proofs of technical results

Proof of Proposition 3.2

Proof. Note that the total number of paths is $P = d_{\text{in}} \cdot (w_{\text{cv}} \cdot w)^{d_{\text{cv}}} \cdot w_{\text{fc}}^{(d_{\text{fc}}-1)}$, and in the definition of NPV for CNNs in Definition 3.3 the indices are over only the weights without specifying the input node given by $\mathcal{I}^{\text{f}}_0(p)$. $\square$



Proof of Proposition 3.3

Proof. For $l=0$, we have $z_{rot(x,r),\Theta}(i_{\text{fout}},1,0)=rot(x,r)(i_{\text{fout}})=x(i_{\text{fout}}\oplus r)=z_{x,\Theta}(i_{\text{fout}}\oplus r,1,0)$. Now for $l=1$, we have

$$\begin{aligned}
q_{rot(x,r),\Theta}(i_{\text{fout}},\cdot,1) &= \sum_{i_{\text{in}}\in[1],i_{\text{cv}}\in[w_{\text{cv}}]}\Theta(i_{\text{cv}},i_{\text{in}},i_{\text{out}},l)\cdot z_{rot(x,r),\Theta}(i_{\text{fout}}\oplus(i_{\text{cv}}-1),i_{\text{in}},0)\\
&= \sum_{i_{\text{in}}\in[1],i_{\text{cv}}\in[w_{\text{cv}}]}\Theta(i_{\text{cv}},i_{\text{in}},i_{\text{out}},l)\cdot z_{x,\Theta}((i_{\text{fout}}\oplus r)\oplus(i_{\text{cv}}-1),i_{\text{in}},0)\\
&= q_{x,\Theta}(i_{\text{fout}}\oplus r,\cdot,1)
\end{aligned}$$

The proof follows by noting that $G=\mathbf{1}_{\{q>0\}}$, and $z=q\cdot G$, and repeating the above argument for the layer $l=2,\ldots,d_{\text{cv}}$. □

Proof of Lemma 4.2

Proof.

$$\begin{aligned}
\langle\phi_{x_s,\Theta},\phi_{x_{s'},\Theta}\rangle &= \sum_{p\in[P]}x_s(\mathcal{I}_0(p))x_{s'}(\mathcal{I}_0(p))A_\Theta(x_s,p)A_\Theta(x_{s'},p)\\
&= \sum_{i=1}^{d_{\text{in}}}x_s(i)x_{s'}(i)\Lambda_\Theta(i,x_s,x_{s'})\\
&= \langle x_s,x_{s'}\rangle_{\Lambda_\Theta(\cdot,x_s,x_{s'})} \qquad (1)
\end{aligned}$$

Owing to the symmetry in a fully connected network, we have $\Lambda(i,x_s,x_{s'})$ to be the same for all values of $i\in[d_{\text{in}}]$. And since $H^{\text{lyr}}_{l,\Theta}(s,s')$ measure the number of gates in layer 'l' that are active for both inputs x_s and $x_{s'}$, the total number of paths active for both inputs is $\Pi_{l=1}^{(d-1)}H^{\text{lyr}}_{l,\Theta}(s,s')$. □

Proof of Lemma 4.2

Proof. Proof is complete by noting that the NPF of the ResNet is a concatenation of the NPFs of the 2^b distinct sub-FC-DNNs within the ResNet architecture. □

Proof of Lemma 4.3

Proof. For the CNN architecture considered in this paper, each bundle has exactly d_{in} number of paths, each one corresponding to a distinct input node. For a bundle $b_{\hat{p}}$, let $b_{\hat{p}}(i), i \in [d_{\text{in}}]$ denote the path starting from input node i.

$$
\begin{aligned}
&\sum_{\hat{p}\in[\hat{P}]}\left(\sum_{i,i'\in[d_{\text{in}}]} x(i)x'(i')A_\Theta\left(x, b_{\hat{p}}(i)\right)A_\Theta\left(x', b_{\hat{p}}(i')\right)\right)\\
&=\sum_{\hat{p}\in[\hat{P}]}\left(\sum_{i\in[d_{\text{in}}],i'=i\oplus r,r\in\{0,\ldots,d_{\text{in}}-1\}} x(i)x'(i\oplus r)A_\Theta\left(x, b_{\hat{p}}(i)\right)A_\Theta\left(x', b_{\hat{p}}(i\oplus r)\right)\right)\\
&=\sum_{\hat{p}\in[\hat{P}]}\left(\sum_{i\in[d_{\text{in}}],r\in\{0,\ldots,d_{\text{in}}-1\}} x(i)rot(x',r)(i)A_\Theta\left(x, b_{\hat{p}}(i)\right)A_\Theta\left(rot(x',r), b_{\hat{p}}(i)\right)\right)\\
&=\sum_{r=0}^{d_{\text{in}}-1}\left(\sum_{i\in[d_{\text{in}}]} x(i)rot(x',r)(i)\sum_{\hat{p}\in[\hat{P}]}A_\Theta\left(x, b_{\hat{p}}(i)\right)A_\Theta\left(rot(x',r), b_{\hat{p}}(i)\right)\right)\\
&=\sum_{r=0}^{d_{\text{in}}-1}\left(\sum_{i\in[d_{\text{in}}]} x(i)rot(x',r)(i)\Lambda_\Theta(i,x,rot(x',r))\right)\\
&=\sum_{r=0}^{d_{\text{in}}-1}\langle x, rot(x',r)\rangle_{\Lambda_\Theta(\cdot,x,rot(x',r))}
\end{aligned}
$$

□

Proof of Theorem 5.1 follows in the same manner as the proof Theorem 5.1 of Lakshminarayanan and Singh (2020).