Efficient Semantic Segmentation using Gradual Grouping
Nikitha Vallurupalli, SriHarsha Annamaneni, Girish Varma, C V Jawahar, Manu Mathew, Soyeb Nagori
IIIT Hyderabad, Texas Instruments Bangalore

REAL TIME SEMANTIC SEGMENTATION
- Consume energy efficiently (Portability)
- Give real-time output (30fps)
- in constrained memory and
- High accuracy for safety
- Cloud not an option
- Large latencies for real-time output
- Violates user privacy.
- Consumes network bandwidth.

RECENT TECHNIQUES TO MAKE CNN’S EFFICIENT
- Depthwise separable convolutions
- Grouped Convolutions
- Shuffled Convolutions

PROPOSED TRAINING METHOD
- GRADUAL GROUPING

- Training procedure where the train time optimization happens in the higher dimensional space of dense convolutions and gradually evolves towards grouped convolutions.
- Start with a dense convolution and multiply the blue edges by a parameter $\alpha$.
- Decrease $\alpha$ gradually during training time from 1 and by the end of the training $\alpha$ becomes 0.
- In fine tuning phase, $\alpha$ remains 0. Finally at test time, the convolutions can be implemented as grouped convolutions which gives better efficiency.

MODEL COMPRESSION

Previous Approaches

Quantization
- Performance improves with quantization
- $W_\text{Q}=\frac{1}{2}$ if $W_\text{Q} \geq 0$, $W_\text{Q}=0$, $W_\text{Q} < 0$.
- High precision arithmetic essential for obtaining high performance.
- This results in memory savings and faster computation.

- Sparsify Quantized Full Frame CNN for Low Power Embedded Devices, One Presentation at CVPRW’17, Manu Mathew, Kumar Deesappa, Pradip Roy, Soyeb Nagori

MICRO LEVEL ARCHITECTURE MODULES

- Non-bottleneck layer used in ERFNet. D, DSC and and DDGS are our proposed micro level layer architectures.
- We have experimented with these proposed micro level layers on the ERF Net baseline micro architecture and studied the effect of changing each module in the encoder. Decoder is not yet optimized.

RESULTS
- Our prime focus was on obtaining compressed models with < 20 GFLOPs and with minimal loss in accuracy.
- We pretrain our proposed encoder on imagenet dataset using gradual grouping, and then attach the light weight decoder to it.
- Selective application of groups (green boxes) hardly degrades the accuracy while still giving a reasonable reduction in GFLOPs of 1.5x over the baseline ERFNet which runs at 27.7 GFLOPs.

- Our method gives a 5X reduction in FLOPs with only 4% degradation in accuracy.
- Blue points representing models trained by gradual grouping gives the best performance tradeoffs.