

進捗報告

表 1: 実験の設定

model	VGG19
Optim(model)	SGD(lr=0.01, momentum=0.9)
Optim(α)	Adam(lr=0.005, $\beta=(0.5, 0.999)$)
Loss	Cross Entropy Loss
dataset	cifar10
batch size	64
train data	25000 + 25000 / 25000
epoch	50 / 100

表 2: Accuracy の比較

	accuracy(%)	学習時間	データ数
探索	88.24	50	50000
評価	87.96	100	25000
ベースライン	87.96	100	25000

1 問題

ショートカットの数を固定していたため、ベースラインが探索空間に入っていなかった。そこでショートカット数も学習に加えた。

ブロックの定義にショートカット数 β を加え、

$$x_i = \text{ConvBn}_i(x_{i-1}) + \beta_i * \sum_j \alpha_{ij} * \text{shortcut}_{ij}(x_j) \quad (1)$$

のように変えた。そのまま学習すると、 $\beta = 0$ で勾配の更新ができなくなるので、

$$\hat{\beta} = \begin{cases} \exp(\beta - 1) & (\beta \leq 1) \\ \log(\beta) + 1 & (\text{otherwise}) \end{cases} \quad (2)$$

指数関数で値域を 0 より大きくした $\hat{\beta}$ (初期値 1) を使用する。log は特に理由はない

2 実験

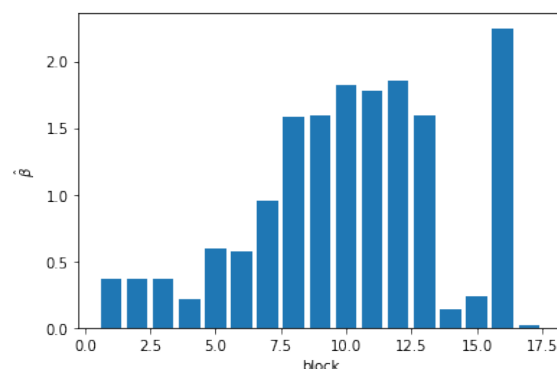
表 1 に実験設定を示した。ショートカット関数は stride ありのとき、Factorized Reduce を使う。

評価段階では $\hat{\beta}$ の小数点を丸めて整数にし、 α の上位から順にショートカットを選んでグラフとした。

(評価段階のデータサイズを間違えて 25000 で学習してしまった。)

2.1 結果

表 2 に結果を示した。ベースラインは下回らなくなった。さらに精度を高めるには、1 としているショート

図 1: 探索後の $\hat{\beta}$ の値

カットの重みが探索時のものを引き継ぐ形式が考えられる。

図 1 には、ブロックごとの $\hat{\beta}$ の重みを、図 2 は得られたグラフを示した。図 3 の初期状態から複数本とるか、取らないかなどの選択が行えていることが分かる。Block 13, 14 の重みが少ないのは、重みの大きい Block 15 から参照されているため、Block 12 とのバラエティを保っていると考えられる。

3 今後の予定

- ショートカット関数の改善
- β 周りの改良

4 ソースコード

github の notebook リポジトリ参照

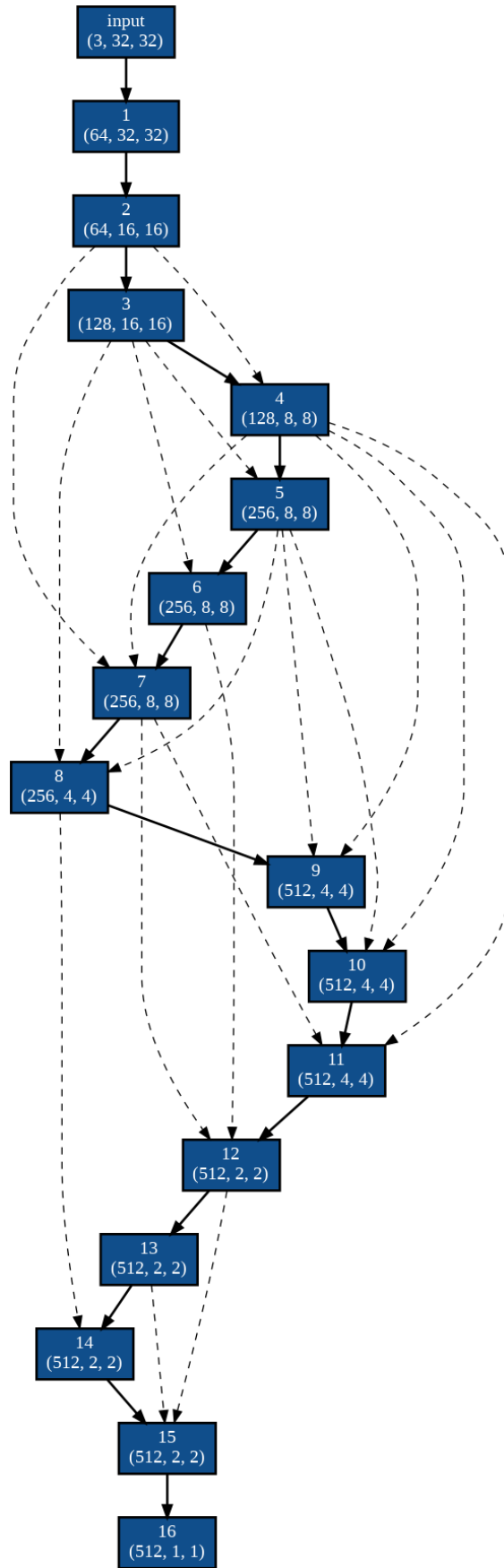


図 2: 探索したグラフ (50 epoch)

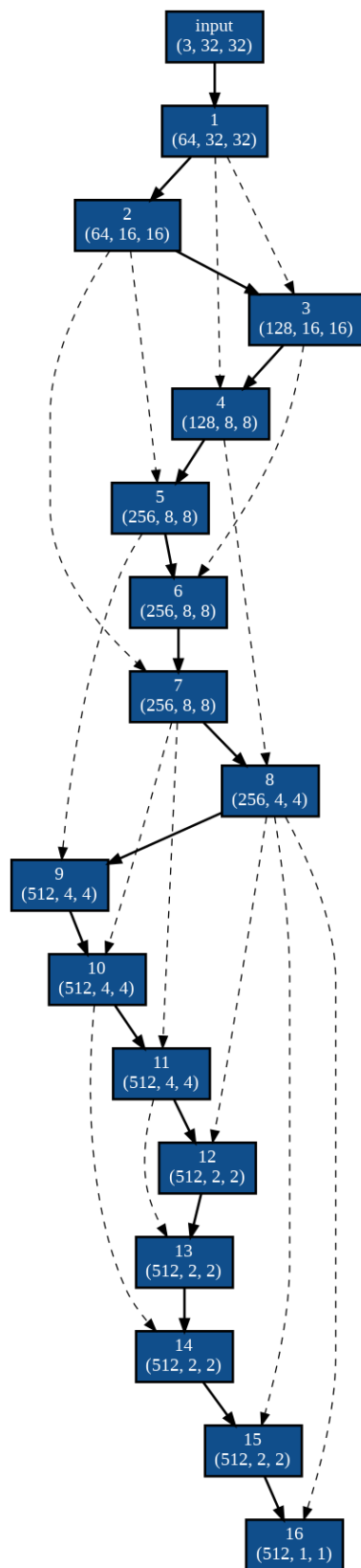


図 3: 未探索のグラフ (0 epoch)