

進捗報告

表 1: 実験の設定

base model	VGG19
Optim(w)	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

1 問題

前回同様.

初期条件として探索をベースラインの VGG19 に相当する α から始めた.

2 実験

表 1 に探索時の実験設定を示した.

評価段階では, 表 1 に加え, SGD の学習率を, 指数スケジューラ ($\gamma = 0.9261 : \gamma^{30} = 0.1$) で減衰させた.

- (a) **descending** α_j の上位から順に, $\text{round}(\hat{\beta}_j)$ 本のショートカットを選んだグラフ,
- (b) **threshold** $\hat{\alpha}_{ij} \geq 0.5$ となるショートカットを選んだグラフ,
- (c) **baseline** ショートカットをすべて破棄したグラフ,
- の 3 つで性能を評価した.

2.1 結果

図 1 にテスト精度を, 表 2 に結果を示した.

図 2 は探索した隣接行列の重み $\hat{\alpha}_{ij} = \hat{\beta}_j * \alpha_{ij}$ である.

図 3, 4, 5 は, 図 2 からグラフ構築手法 (a)(b)(c) で得られたグラフである. ノードが各ブロックの出力を示し, エッジは太線が VGG のレイヤー, 点線がショートカットを表している.

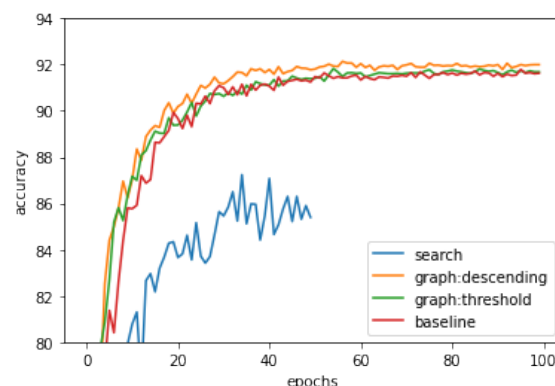


図 1: 学習中のテスト精度

図 2: 探索後の隣接行列の重み $\hat{\alpha}$

3 考察

予備実験で, (a) の実験を Step スケジューラ, 300 epoch で行ったところ, テスト精度が 93.97 %まで伸びたので, 時間があればこの設定で学習したい.

4 今後の予定

- 探索 10 回試行したグラフの比較

5 ソースコード

github の notebook リポジトリ参照. 探索中のグラフの変化は, この資料の ./graph を参照.

表 2: 各条件の比較

	テスト精度 (%)	学習時間 (epoch)	計算時間 (GPU-min)	パラメータ数	データサイズ		
					train	valid	test
探索	87.24	50	120	26.30M	25000	25000	5000
(a) descending	92.14	100	100	20.85M	50000	0	10000
(b) threshold	91.83	100	100	20.27M	50000	0	10000
(c) baseline	91.79	100	100	20.04M	50000	0	10000

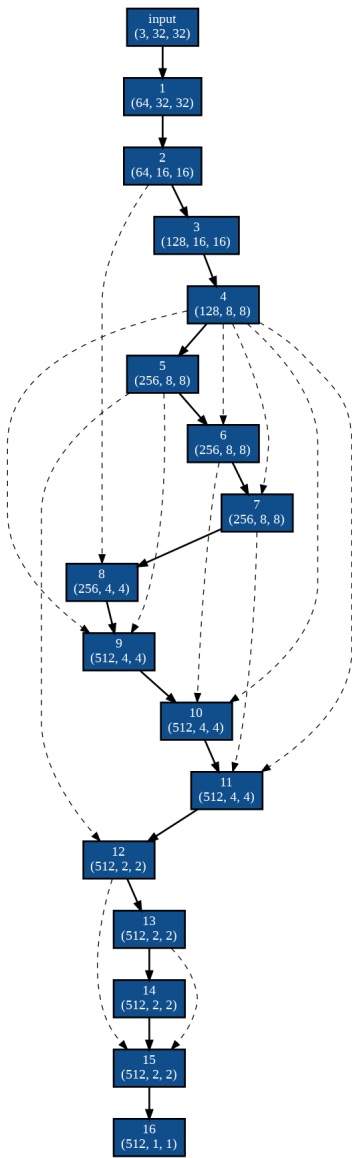


図 3: グラフ (a)

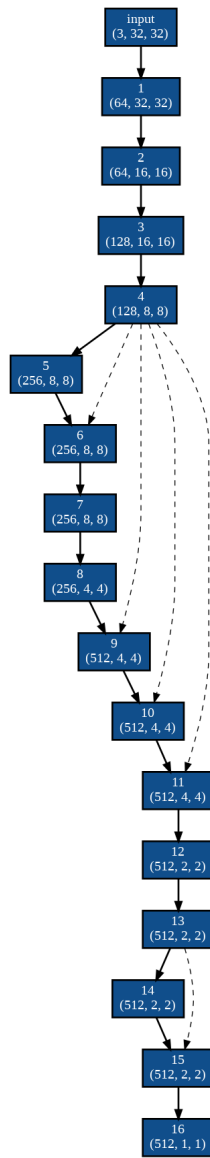


図 4: グラフ (b)

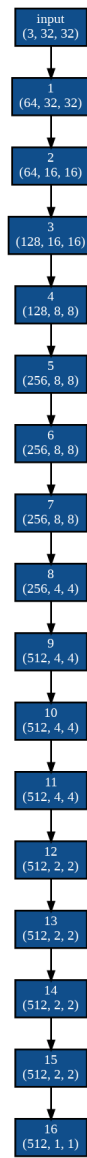


図 5: グラフ (c)