

進捗報告

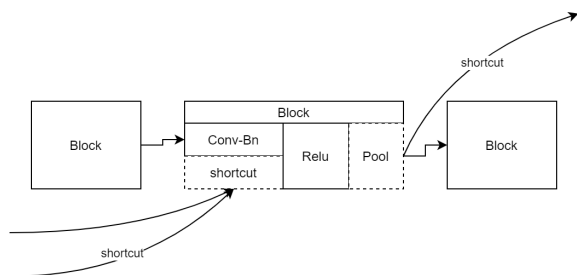


図 1: ショートカット位置

表 1: 実験の設定

model	VGG11
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
epoch	100

1 今週やったこと

- VGG のアーキテクチャ探索の実装と実験

2 問題設定

DARTS の追実験が概ね成功したので, 新たにアーキテクチャの探索を設定してみることにした. 調査の結果, VGG の最適なショートカット位置を DARTS で探索することを問題に定めた.

図 1 にショートカット位置の関係を示した.

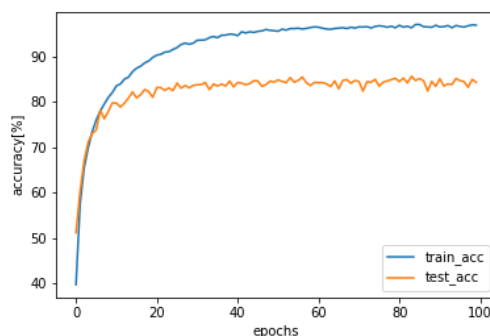


図 2: 精度

3 実験

ショートカット部分は

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

のように, concat ではなくショートカットの加重和との和とした.

表 1 に実験設定を示した. もととなるモデルは VGG11 とした. したがって Block 数は 8 となった. この Block に対して可能な接続の組み合わせを全て探索することにした.

今回はアーキテクチャ探索 (α を得る) 段階までを行い, 性能は評価していない.

4 結果

図 2 には訓練とテストの精度を示した. テスト精度は最大で 85.54%であった.

表 2, 3 にはアーキテクチャー探索結果の α の重みを示した.

5 考察

VGG に対してショートカットの重みを学習できることを確認した.

DARTS は得られた重みに対して, 適切なアーキテクチャを決定する必要がある. 決定する方法はいくつか考えられるため, 次回はそれぞれの手法の精度を比較したい. 図 3 には必ず 1 つの接続を持つことを仮定した場合の VGG の構造を示した.

6 今後の予定

- アーキテクチャの性能評価

表 2: α の重み (softmax)

$i \setminus j$	1	2	3	4	5	6
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.2019	0.7981	0.0000	0.0000	0.0000	0.0000
5	0.2144	0.1256	0.6600	0.0000	0.0000	0.0000
6	0.0588	0.8246	0.0570	0.0596	0.0000	0.0000
7	0.0368	0.8497	0.0374	0.0386	0.0375	0.0000
8	0.1238	0.1484	0.2058	0.2357	0.2417	0.0446

表 3: α の重み

$i \setminus j$	1	2	3	4	5	6
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	-0.6871	0.6871	0.0000	0.0000	0.0000	0.0000
5	-0.1982	-0.7329	0.9261	0.0000	0.0000	0.0000
6	-0.6557	1.9855	-0.6870	-0.6411	0.0000	0.0000
7	-0.6814	2.4581	-0.6656	-0.6325	-0.6629	0.0000
8	-0.1647	0.0171	0.3440	0.4793	0.5046	-1.1853

7 ソースコード

Github の同階層を参照.

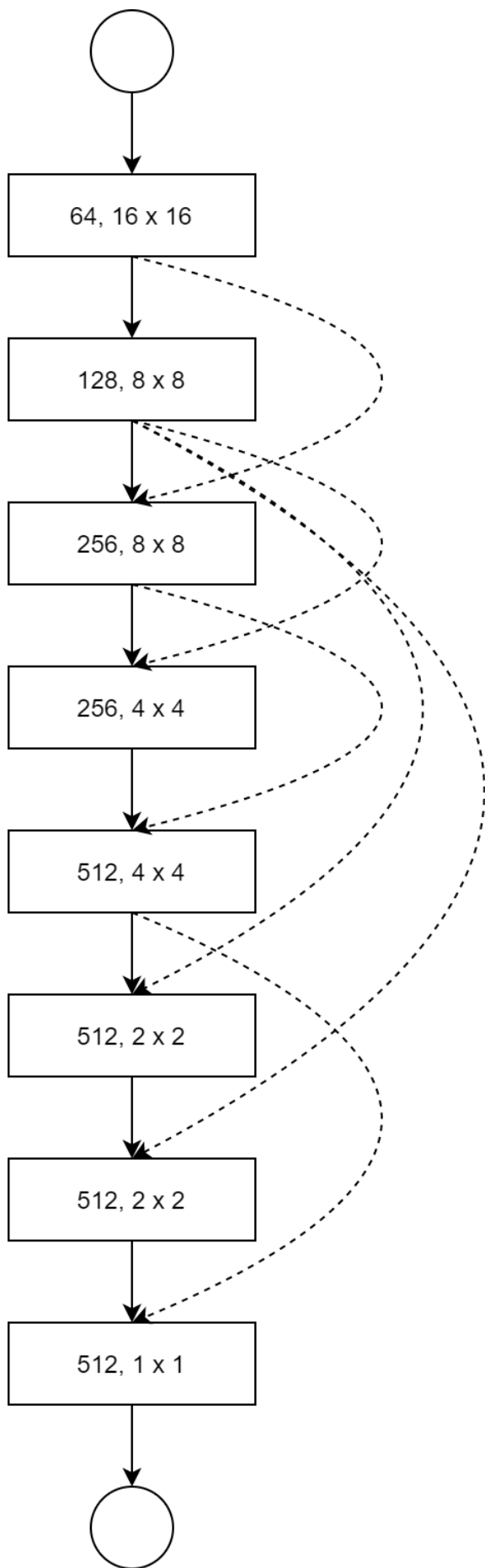


図 3: モデル