

進捗報告

表 1: モデルの設定

base model	VGG19
Optim(w)	SGD(lr=0.001, momentum=0.9)
Optim(α)	Adam(lr=0.001, $\beta=(0.5, 0.999)$)
Loss	Cross Entropy Loss
dataset	cifar10
pretrain	true
batch size	64
train size	12500
valid size	5000

表 2: GA の設定

個体数	10
世代数	60
選択	TD 選択
温度	10 \rightarrow ??
交叉	一様交叉
交叉率	0.5 (0.5)
変異	ガウス分布
変異率	0.2 (0.2)

1 今週やったこと

- TDGA の実装

2 変更

以前までブロック単位で Softmax していた (1) 式を, 辺単位で正規化する (2) 式に変更してみた.

$$x_i = f_{i-1,i}^c(x_{i-1}) + \beta_i \sum_{j \in S_i} \text{Softmax}(\alpha_i)_j f_{j,i}^s(x_j) \quad (1)$$

$$x_i = f_{i-1,i}^c(x_{i-1}) + \sum_{j \in S_i} \text{Sigmoid}(\alpha_{ij}) f_{j,i}^s(x_j) \quad (2)$$

これによって, β の補正がなくなり, 3 本のショートカットができたり, 全体の本数が増えたりした.

3 実験

表 1, 2 にモデルと GA の設定を示した.

(温度設定は, 10 \rightarrow 2 \rightarrow 1 にしたつもりが, 始点と終点がぐちゃぐちゃになっていた.)

4 結果

図 1 は最終世代の最良個体のグラフ, アーキテクチャの評価は 93.68 %

図 4 はショートカット数の平均だが, 分散を保ちつつ減少方向へ学習できていることが分かる.

図 3 はテスト accuracy の平均を示す. GA なしの実験時は 88 % 程度だったのが, 82 % 程度で伸びが悪い. One Shot モデルで w を共有しているため, それぞれの個体に引っ張られて学習が遅いと思われる.

5 今後の予定

- 動作は確認できたので, 世代数を増やすなどのパラメータの見直しを検討.
- 卒論の準備.

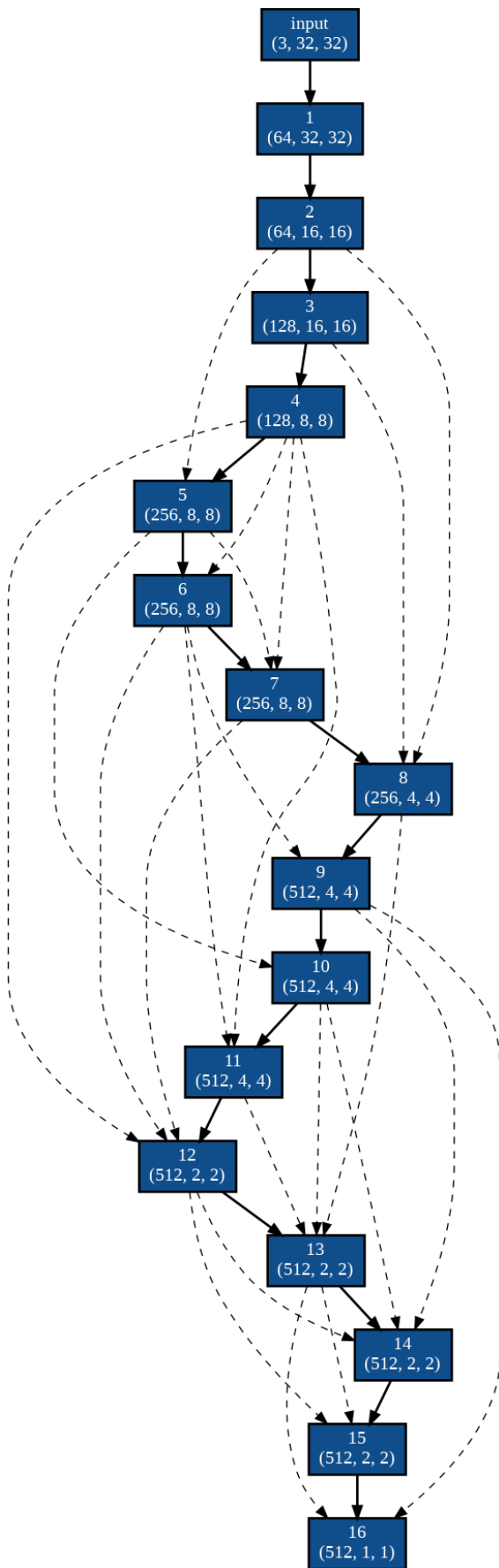


図 1: 最終世代の最良個体のグラフ

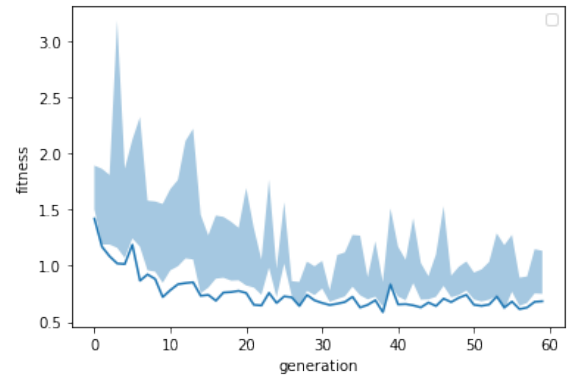


図 2: 世代ごとの fitness

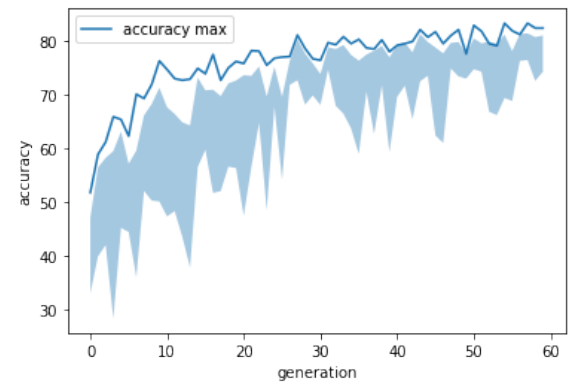


図 3: 世代ごとの test accuracy

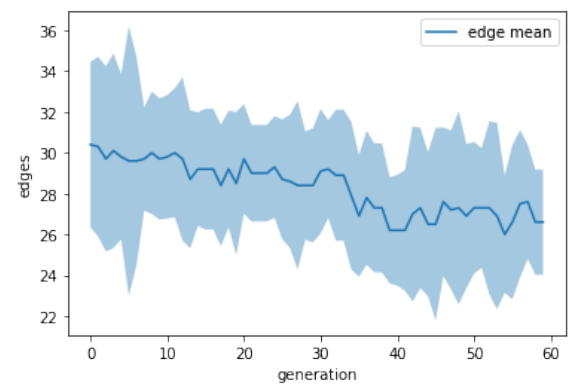


図 4: 世代ごとのショートカット数