Московский государственный университет имени М. В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра математических методов прогнозирования

Васильев Руслан Леонидович

# Калибровка уверенности нейросетей

КУРСОВАЯ РАБОТА

**Научный руководитель:**
д.ф-м.н., профессор
*А. Г. Дьяконов*

Москва, 2021

# Содержание

**Аннотация**

Аннотация обычно содержит краткое описание постановки задачи и полученных результатов, одним абзацем на 10–15 строк. Цель аннотации — обозначить в общих чертах, о чём работа, чтобы человек, совершенно не знакомый с данной работой, понял, интересна ли ему эта тема, и стоит ли читать дальше. Аннотация собирается в последнюю очередь путем легкой модификации наиболее важных и удачных фраз из введения и заключения.

# 1 Классификация, откалиброванность

Пусть решается задача классификации объектов из множества $\mathcal{X}$ с метками (классами) $\mathcal{Y} = \{1, \ldots, n\}$. Предположим, что с помощью обучающей выборки – множества пар объектов и соответствующих им меток $(x_i, y_i)_{i=1}^{l}$ – мы обучили модель – алгоритм, для каждого $x \in \mathcal{X}$ выдающую вектор оценок – *уверенностей* (confidences) $\mathbf{a}(x) = (a_1(x), \ldots, a_n(x))$, $\sum_{j=1}^{n} a_j(x) = 1$. Далее объекту приписывается класс, соответствующий наибольшей уверенности:

$$\hat{y}(x) := \operatorname*{argmax}_{j \in \mathcal{Y}} a_j, \quad \hat{p}(x) := a_{\hat{y}}. \tag{1}$$

Оценку $\hat{p}$ мы бы хотели трактовать как вероятность того, истинная метка $y$ совпадает с предсказанной $\hat{y}$. Если наша оценка достаточно точна, то модель называют *откалиброванной*. Например, если мы рассматриваем объекты для каждого из которых $\hat{p} \approx 0.8$, то мы ожидаем, что $\approx 80\%$ из них будут классифицированы верно. Формально определение *откалиброванности* (в [1] – perfect calibration) можно записать следующим образом:

$$\mathbb{P}\left(y = \hat{y} \mid \hat{p} = p\right) = p \quad \forall p \in [0, 1]\,. \tag{2}$$

В случае реальных данных и моделей мы не можем проверить (2), поэтому на помощь приходят различные метрики и визулизации, которые будут рассмотрены в разделе <…>.

Существуют и более сильные определения откалиброванности модели, чем (2). Например, согласно [2] классификатор называется откалиброванным (в оригинале – well-calibrated), если

$$\mathbb{P}(y = j \mid a_j = p) = p \quad \forall j \in \mathcal{Y}, \quad \forall p \in [0, 1]\,, \tag{3}$$

то есть мы ожидаем, что уверенности, выдываемые для каждого класса (а не только предсказанного), являются откалиброванными. Еще более сильно откалиброванность определяется в [3]:

$$\mathbb{P}(y = j \mid \mathbf{a} = \mathbf{p}) = p_j \quad \forall j \in \mathcal{Y}, \quad \forall \mathbf{p} \in \Delta^{n-1}, \tag{4}$$

где $\Delta^{n-1} = \left\{\mathbf{p} \in [0, 1] : \sum_{j=1}^{n} p_j = 1\right\}$.

<…>Может, сюда вставить постановку задачи – сказать, что мы хотим преобразовать выходы модели?..

# 2 Как оценить откалиброванность

## 2.1 Метрики

Одна из наиболее популярных метрик для оценки откалиброванности модели – ECE. MCE. Class-wise ECE. Brier. NLL.

## 2.2 Визуализация

# 3 Методы калибровки

## 3.1 Гистограммный биннинг (Histogram Binning)

## 3.2 Изотоническая регрессия

## 3.3 Линейные отображения логитов

## 3.4 Сглаживание меток (Label Smoothing)

## 3.5 А также...

# 4 Вычислительные эксперименты

Эксперименты были проведены с архитектурами <. . . >на датасетах CIFAR-10, CIFAR-100, Imagenet. При вычислениях были использованы предобученные модели из открытых репозиториев [4, 5, 6]. На выходе

на то, что эфишент нет не переуверенна обратили внимание еще в статье. причина - лаплас.

# 5 Почему нейросети не откалиброваны?

# 6 Заключение

# 7 Приложения

# 8 Список литературы

# Список литературы

[1] Chuan Guo et al. “On Calibration of Modern Neural Networks”. In: *ICML 2017*. Vol. 70. Proceedings of Machine Learning Research. PMLR, 2017, pp. 1321–1330.

*Таблица 1: Accuracy, % – доля правильных ответов (больше – лучше). Значения метрики приводятся для тестовой выборки до и после калибровки.*

| Данные | Модель | До калибровки | Hist-binning | Isotonic | T-scaling | V-scaling | V-scaling + bias |
|---|---|---|---|---|---|---|---|
| cifar10 | densenet121 | **93.960** | 93.680 | 93.800 | **93.960** | 93.900 | 93.860 |
| cifar10 | densenet161 | **94.040** | 93.580 | 93.860 | **94.040** | 93.920 | **94.040** |
| cifar10 | densenet169 | **94.400** | 94.160 | 94.220 | **94.400** | 94.280 | 94.240 |
| cifar10 | googlenet | **93.040** | 92.700 | 92.900 | **93.040** | 93.000 | 93.020 |
| cifar10 | inception_v3 | 93.380 | 93.280 | 93.320 | 93.380 | **93.420** | 93.360 |
| cifar10 | mobilenet_v2 | **93.180** | 92.920 | 92.960 | **93.180** | 93.060 | 93.040 |
| cifar10 | resnet18 | 92.960 | 92.840 | **93.140** | 92.960 | 93.020 | 93.040 |
| cifar10 | resnet34 | **93.420** | 93.020 | 93.180 | **93.420** | 93.380 | 93.340 |
| cifar10 | resnet50 | **93.580** | 93.400 | 93.520 | **93.580** | **93.580** | 93.560 |
| cifar10 | vgg11_bn | **92.200** | 91.800 | 91.880 | **92.200** | 91.980 | 92.060 |
| cifar10 | vgg13_bn | 93.980 | 93.680 | 93.800 | 93.980 | **94.080** | 93.980 |
| cifar10 | vgg16_bn | **93.880** | 93.560 | 93.600 | **93.880** | 93.720 | 93.760 |
| cifar10 | vgg19_bn | 93.680 | 93.460 | 93.620 | 93.680 | 93.580 | **93.700** |
| cifar100 | mobilenetv2_x0_5 | **71.720** | 68.520 | 71.400 | **71.720** | 71.220 | 71.420 |
| cifar100 | mobilenetv2_x1_0 | 74.760 | 72.440 | 74.260 | 74.760 | **74.820** | 74.580 |
| cifar100 | mobilenetv2_x1_4 | **76.120** | 74.040 | 75.400 | **76.120** | 76.020 | **76.120** |
| cifar100 | resnet20 | **68.680** | 65.300 | 67.800 | **68.680** | 68.540 | 68.320 |
| cifar100 | resnet32 | **70.120** | 67.120 | 69.420 | **70.120** | 69.620 | 69.560 |
| cifar100 | resnet44 | **71.860** | 69.060 | 71.300 | **71.860** | 71.520 | 71.320 |
| cifar100 | resnet56 | **73.140** | 70.840 | 72.660 | **73.140** | 72.920 | 72.760 |
| cifar100 | shufflenetv2_x0_5 | 67.660 | 65.220 | 67.920 | 67.660 | **68.060** | **68.060** |
| cifar100 | shufflenetv2_x1_0 | 72.840 | 70.760 | 72.560 | 72.840 | **73.220** | 72.960 |
| cifar100 | shufflenetv2_x1_5 | 74.440 | 71.780 | 74.140 | 74.440 | **74.520** | **74.520** |
| cifar100 | shufflenetv2_x2_0 | **75.660** | 73.840 | 75.180 | **75.660** | 75.420 | 75.440 |
| cifar100 | vgg11_bn | **70.540** | 68.740 | 70.380 | **70.540** | 70.360 | 70.340 |
| cifar100 | vgg13_bn | **74.320** | 72.200 | 73.480 | **74.320** | 74.180 | 73.880 |
| cifar100 | vgg16_bn | **74.000** | 72.420 | 73.680 | **74.000** | 73.840 | 73.780 |
| cifar100 | vgg19_bn | 74.000 | 72.720 | **74.080** | 74.000 | 73.980 | 73.860 |
| imagenet | mobilenetv2_120d | **77.220** | 74.000 | 76.528 | **77.220** | 77.188 | 77.060 |
| imagenet | repvgg_b3 | **80.320** | 77.464 | 79.820 | **80.320** | 80.240 | 80.236 |
| imagenet | tf_efficientnet_b8 | 85.428 | 83.756 | 85.232 | 85.428 | 85.420 | **85.440** |
| imagenet | vgg19_bn | 74.140 | 70.920 | 73.680 | 74.140 | **74.172** | 73.768 |

[2] Bianca Zadrozny and Charles Elkan. “Transforming Classifier Scores into Accurate Multiclass Probability Estimates”. In: *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’02. Edmonton, Alberta, Canada: Association for Computing Machinery, 2002, pp. 694–699. ISBN: 158113567X. DOI: 10.1145/775047.775151. URL: https://doi.org/10.1145/775047.775151.

[3] Meelis Kull et al. “Beyond temperature scaling: Obtaining well-calibrated multi-class probabilities with Dirichlet calibration”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc., 2019. URL: https://proceedings.neurips.cc/paper/2019/file/8ca01ea920679a0fe3728441494041b9-Paper.pdf.

*Таблица 2: Brier Score (меньше – лучше). Значения метрики приводятся для тестовой выборки до и после калибровки.*

| Данные | Модель | До калибровки | Hist-binning | Isotonic | T-scaling | V-scaling | V-scaling + bias |
|---|---|---|---|---|---|---|---|
| cifar10 | densenet121 | 0.101 | 0.106 | **0.098** | 0.102 | 0.102 | 0.102 |
| cifar10 | densenet161 | 0.099 | 0.105 | **0.095** | 0.099 | 0.098 | 0.098 |
| cifar10 | densenet169 | 0.093 | 0.097 | **0.089** | 0.093 | 0.093 | 0.092 |
| cifar10 | googlenet | 0.108 | 0.113 | 0.108 | 0.108 | **0.107** | 0.107 |
| cifar10 | inception_v3 | 0.105 | 0.113 | **0.103** | 0.106 | 0.106 | 0.105 |
| cifar10 | mobilenet_v2 | 0.103 | 0.113 | **0.101** | 0.104 | 0.105 | 0.105 |
| cifar10 | resnet18 | 0.110 | 0.114 | **0.108** | 0.109 | 0.109 | 0.108 |
| cifar10 | resnet34 | 0.109 | 0.116 | **0.104** | 0.107 | 0.107 | 0.106 |
| cifar10 | resnet50 | 0.103 | 0.107 | **0.098** | 0.102 | 0.102 | 0.102 |
| cifar10 | vgg11_bn | 0.118 | 0.125 | **0.117** | 0.117 | 0.118 | 0.118 |
| cifar10 | vgg13_bn | 0.091 | 0.101 | **0.091** | 0.092 | 0.091 | 0.091 |
| cifar10 | vgg16_bn | 0.098 | 0.105 | **0.095** | 0.097 | 0.097 | 0.097 |
| cifar10 | vgg19_bn | 0.102 | 0.108 | **0.098** | 0.101 | 0.101 | 0.100 |
| cifar100 | mobilenetv2_x0_5 | 0.415 | 0.450 | 0.398 | 0.393 | **0.393** | 0.393 |
| cifar100 | mobilenetv2_x1_0 | 0.372 | 0.408 | 0.360 | 0.354 | 0.353 | **0.353** |
| cifar100 | mobilenetv2_x1_4 | 0.354 | 0.389 | 0.344 | 0.339 | **0.338** | 0.338 |
| cifar100 | resnet20 | 0.452 | 0.488 | 0.441 | 0.432 | **0.432** | 0.434 |
| cifar100 | resnet32 | 0.444 | 0.475 | 0.421 | 0.412 | **0.412** | 0.413 |
| cifar100 | resnet44 | 0.424 | 0.456 | 0.398 | **0.391** | 0.391 | 0.392 |
| cifar100 | resnet56 | 0.414 | 0.434 | 0.384 | **0.378** | 0.379 | 0.380 |
| cifar100 | shufflenetv2_x0_5 | 0.458 | 0.493 | 0.439 | **0.433** | 0.434 | 0.436 |
| cifar100 | shufflenetv2_x1_0 | 0.397 | 0.433 | 0.384 | **0.379** | 0.380 | 0.380 |
| cifar100 | shufflenetv2_x1_5 | 0.372 | 0.413 | 0.365 | **0.362** | 0.364 | 0.365 |
| cifar100 | shufflenetv2_x2_0 | 0.350 | 0.386 | 0.345 | **0.344** | 0.345 | 0.345 |
| cifar100 | vgg11_bn | 0.445 | 0.458 | 0.413 | **0.407** | 0.409 | 0.409 |
| cifar100 | vgg13_bn | 0.401 | 0.421 | 0.378 | **0.372** | 0.374 | 0.373 |
| cifar100 | vgg16_bn | 0.439 | 0.432 | 0.376 | **0.371** | 0.373 | 0.372 |
| cifar100 | vgg19_bn | 0.442 | 0.426 | 0.369 | 0.370 | 0.369 | **0.368** |
| imagenet | mobilenetv2_120d | 0.327 | 0.376 | 0.326 | 0.319 | **0.318** | 0.321 |
| imagenet | repvgg_b3 | 0.286 | 0.333 | 0.289 | 0.286 | **0.284** | 0.287 |
| imagenet | tf_efficientnet_b8 | 0.225 | 0.249 | **0.217** | 0.218 | 0.218 | 0.220 |
| imagenet | vgg19_bn | 0.358 | 0.420 | 0.365 | 0.357 | **0.357** | 0.360 |

[4] Huy Phan. *huyvnphan/PyTorch_CIFAR10*. Version v3.0.1. Jan. 2021. DOI: 10.5281/zenodo.4431043. URL: https://doi.org/10.5281/zenodo.4431043.

[5] chenyaofo. *PyTorch CIFAR models*. 2021. URL: https://github.com/chenyaofo/pytorch-cifar-models.

[6] Ross Wightman. *PyTorch Image Models*. https://github.com/rwightman/pytorch-image-models. 2019. DOI: 10.5281/zenodo.4414861.

*Таблица 3: ECE, % – Expected Calibration Error, 15 бинов (меньше – лучше). Значения метрики приводятся для тестовой выборки до и после калибровки.*

| Данные | Модель | До калибровки | Hist-binning | Isotonic | T-scaling | V-scaling | V-scaling + bias |
|---|---|---|---|---|---|---|---|
| cifar10 | densenet121 | 1.90 | **1.03** | 1.89 | 1.74 | 1.76 | 1.64 |
| cifar10 | densenet161 | 2.09 | 1.54 | **1.47** | 1.92 | 2.09 | 2.03 |
| cifar10 | densenet169 | 2.43 | **1.23** | 1.27 | 2.07 | 1.83 | 1.75 |
| cifar10 | googlenet | 1.70 | 1.10 | 1.37 | 1.07 | **0.99** | 1.09 |
| cifar10 | inception_v3 | 2.09 | **1.05** | 1.85 | 1.49 | 1.55 | 1.41 |
| cifar10 | mobilenet_v2 | 2.87 | **1.98** | 2.10 | 2.08 | 2.25 | 2.16 |
| cifar10 | resnet18 | 1.91 | 1.17 | 1.74 | 1.27 | 1.27 | **1.16** |
| cifar10 | resnet34 | 2.52 | 1.76 | **1.51** | 2.18 | 1.98 | 2.20 |
| cifar10 | resnet50 | 2.34 | 1.66 | **1.28** | 1.82 | 1.85 | 2.04 |
| cifar10 | vgg11_bn | 1.71 | **1.50** | 1.54 | 1.62 | 1.62 | 1.82 |
| cifar10 | vgg13_bn | **0.99** | 1.42 | 1.53 | 1.51 | 1.49 | 1.45 |
| cifar10 | vgg16_bn | 1.67 | 1.56 | **1.30** | 1.55 | 1.63 | 1.71 |
| cifar10 | vgg19_bn | 2.26 | 1.47 | **1.28** | 1.90 | 1.98 | 1.95 |
| cifar100 | mobilenetv2_x0_5 | 11.43 | 8.99 | 4.34 | **2.52** | 3.02 | 3.21 |
| cifar100 | mobilenetv2_x1_0 | 10.97 | 8.51 | 5.03 | 3.33 | 3.29 | **3.29** |
| cifar100 | mobilenetv2_x1_4 | 10.25 | 8.97 | 5.10 | 3.64 | 3.52 | **3.49** |
| cifar100 | resnet20 | 10.67 | 9.09 | 5.18 | **2.79** | 3.15 | 3.27 |
| cifar100 | resnet32 | 13.47 | 10.72 | 5.07 | **1.88** | 2.22 | 2.33 |
| cifar100 | resnet44 | 13.89 | 9.59 | 4.67 | **2.22** | 2.45 | 2.82 |
| cifar100 | resnet56 | 13.87 | 9.00 | 5.02 | 2.79 | **2.62** | 3.26 |
| cifar100 | shufflenetv2_x0_5 | 12.43 | 10.50 | 4.39 | **1.51** | 1.78 | 2.41 |
| cifar100 | shufflenetv2_x1_0 | 10.92 | 8.46 | 5.34 | **3.56** | 4.19 | 3.83 |
| cifar100 | shufflenetv2_x1_5 | 9.08 | 8.65 | 5.44 | 4.81 | 4.72 | **4.69** |
| cifar100 | shufflenetv2_x2_0 | 7.36 | 8.49 | 5.09 | 4.56 | **4.38** | 4.46 |
| cifar100 | vgg11_bn | 15.26 | 10.43 | 6.73 | **4.87** | 5.11 | 5.46 |
| cifar100 | vgg13_bn | 13.60 | 8.25 | 7.42 | **6.20** | 6.58 | 6.41 |
| cifar100 | vgg16_bn | 18.94 | 7.46 | 6.08 | 4.09 | **4.05** | 4.13 |
| cifar100 | vgg19_bn | 19.38 | 6.68 | 4.66 | 4.21 | 3.57 | **3.00** |
| imagenet | mobilenetv2_120d | 6.63 | 6.83 | 2.19 | **1.89** | 2.26 | 3.08 |
| imagenet | repvgg_b3 | **3.11** | 6.61 | 3.46 | 3.73 | 3.91 | 4.63 |
| imagenet | tf_efficientnet_b8 | 8.85 | 4.24 | **2.79** | 3.44 | 4.07 | 4.36 |
| imagenet | vgg19_bn | 3.75 | 8.86 | 3.88 | 1.98 | **1.72** | 2.20 |

*Таблица 4: MCE, % – Maximum Calibration Error, 15 бинов (меньше – лучше). Значения метрики приводятся для тестовой выборки до и после калибровки.*

| Данные | Модель | До калибровки | Hist-binning | Isotonic | T-scaling | V-scaling | V-scaling + bias |
|---|---|---|---|---|---|---|---|
| cifar10 | densenet121 | 41.77 | 38.83 | 26.81 | **25.13** | 75.16 | 32.69 |
| cifar10 | densenet161 | 33.63 | 32.88 | 35.10 | 48.49 | 31.01 | **30.55** |
| cifar10 | densenet169 | 42.49 | 25.20 | **23.90** | 33.11 | 25.80 | 24.63 |
| cifar10 | googlenet | 24.83 | 26.12 | 26.23 | 27.21 | 24.79 | **24.46** |
| cifar10 | inception_v3 | 16.93 | 38.86 | 80.05 | 21.97 | **15.07** | 24.74 |
| cifar10 | mobilenet_v2 | 28.72 | 35.87 | **19.17** | 28.92 | 21.72 | 31.27 |
| cifar10 | resnet18 | **15.72** | 36.55 | 29.31 | 19.87 | 25.48 | 43.70 |
| cifar10 | resnet34 | 25.48 | 59.97 | 81.20 | 22.77 | 20.36 | **19.10** |
| cifar10 | resnet50 | 24.96 | 24.32 | 19.31 | 19.00 | **17.85** | 27.30 |
| cifar10 | vgg11_bn | 23.35 | 75.53 | **11.88** | 23.28 | 23.35 | 14.64 |
| cifar10 | vgg13_bn | **14.13** | 31.21 | 20.61 | 32.52 | 24.63 | 83.67 |
| cifar10 | vgg16_bn | 23.53 | 42.56 | 31.32 | **18.02** | 26.22 | 23.75 |
| cifar10 | vgg19_bn | 25.99 | 29.13 | **17.53** | 21.84 | 23.62 | 23.70 |
| cifar100 | mobilenetv2_x0_5 | 25.38 | 19.68 | 13.29 | 93.50 | **6.66** | 8.12 |
| cifar100 | mobilenetv2_x1_0 | 27.73 | 42.98 | 10.70 | 9.78 | **8.24** | 11.05 |
| cifar100 | mobilenetv2_x1_4 | 54.46 | 24.38 | 12.55 | **6.54** | 7.49 | 8.51 |
| cifar100 | resnet20 | 23.59 | 18.06 | 10.76 | 11.50 | **7.79** | 7.92 |
| cifar100 | resnet32 | 38.45 | 21.16 | 12.59 | 7.12 | **6.58** | 7.16 |
| cifar100 | resnet44 | 29.50 | 21.23 | 14.62 | **7.78** | 11.24 | 10.79 |
| cifar100 | resnet56 | 31.35 | 27.13 | 14.06 | 9.59 | 7.62 | **6.52** |
| cifar100 | shufflenetv2_x0_5 | 24.78 | 20.11 | 10.22 | 6.73 | **6.17** | 11.24 |
| cifar100 | shufflenetv2_x1_0 | 29.20 | 21.62 | 12.35 | **8.74** | 9.15 | 8.76 |
| cifar100 | shufflenetv2_x1_5 | 23.80 | 36.70 | 12.08 | **11.09** | 12.43 | 12.97 |
| cifar100 | shufflenetv2_x2_0 | 17.13 | 22.45 | 12.76 | 10.90 | 10.80 | **9.91** |
| cifar100 | vgg11_bn | 40.22 | 25.83 | 20.32 | 11.79 | **10.80** | 10.89 |
| cifar100 | vgg13_bn | 32.80 | 27.00 | 20.77 | **15.01** | 19.07 | 16.03 |
| cifar100 | vgg16_bn | 51.02 | 37.50 | 20.15 | 16.16 | 12.49 | **11.77** |
| cifar100 | vgg19_bn | 50.32 | 36.75 | 16.59 | 28.29 | 28.09 | **11.41** |
| imagenet | mobilenetv2_120d | 12.35 | 14.01 | **5.80** | 5.98 | 14.11 | 10.77 |
| imagenet | repvgg_b3 | 10.41 | 25.70 | **6.68** | 8.73 | 8.63 | 11.89 |
| imagenet | tf_efficientnet_b8 | **11.45** | 25.67 | 11.66 | 13.05 | 12.15 | 13.33 |
| imagenet | vgg19_bn | 8.03 | 19.92 | 7.82 | 7.57 | 7.66 | **4.27** |

*Таблица 5: Negative Log-Likelihood (меньше – лучше). Значения метрики приводятся для тестовой выборки до и после калибровки.*

| Данные | Модель | До калибровки | Hist-binning | Isotonic | T-scaling | V-scaling | V-scaling + bias |
|---|---|---|---|---|---|---|---|
| cifar10 | densenet121 | **0.253** | 0.453 | 0.305 | 0.253 | 0.254 | 0.254 |
| cifar10 | densenet161 | 0.253 | 0.402 | 0.281 | 0.253 | 0.245 | **0.244** |
| cifar10 | densenet169 | 0.228 | 0.384 | 0.243 | 0.227 | 0.224 | **0.224** |
| cifar10 | googlenet | 0.243 | 0.302 | 0.265 | 0.236 | 0.233 | **0.232** |
| cifar10 | inception_v3 | 0.254 | 0.565 | 0.311 | 0.254 | 0.254 | **0.253** |
| cifar10 | mobilenet_v2 | 0.241 | 0.564 | 0.257 | **0.239** | 0.243 | 0.243 |
| cifar10 | resnet18 | 0.256 | 0.407 | 0.334 | 0.255 | 0.253 | **0.253** |
| cifar10 | resnet34 | 0.259 | 0.484 | 0.285 | 0.256 | 0.253 | **0.253** |
| cifar10 | resnet50 | 0.242 | 0.450 | 0.305 | 0.240 | **0.239** | 0.239 |
| cifar10 | vgg11_bn | 0.255 | 0.415 | 0.330 | 0.255 | 0.256 | **0.254** |
| cifar10 | vgg13_bn | 0.206 | 0.430 | 0.339 | 0.206 | 0.205 | **0.205** |
| cifar10 | vgg16_bn | 0.227 | 0.413 | 0.322 | **0.227** | 0.227 | 0.228 |
| cifar10 | vgg19_bn | 0.246 | 0.476 | 0.310 | **0.244** | 0.244 | 0.244 |
| cifar100 | mobilenetv2_x0_5 | 1.163 | 3.666 | 1.505 | **1.033** | 1.033 | 1.034 |
| cifar100 | mobilenetv2_x1_0 | 1.072 | 3.578 | 1.531 | 0.954 | 0.953 | **0.947** |
| cifar100 | mobilenetv2_x1_4 | 1.009 | 3.086 | 1.524 | 0.912 | 0.914 | **0.910** |
| cifar100 | resnet20 | 1.234 | 3.622 | 1.769 | 1.128 | **1.126** | 1.132 |
| cifar100 | resnet32 | 1.328 | 3.818 | 1.560 | 1.117 | 1.115 | **1.114** |
| cifar100 | resnet44 | 1.295 | 3.893 | 1.527 | 1.059 | 1.061 | **1.058** |
| cifar100 | resnet56 | 1.285 | 3.291 | 1.591 | **1.033** | 1.038 | 1.033 |
| cifar100 | shufflenetv2_x0_5 | 1.296 | 3.551 | 1.602 | **1.162** | 1.165 | 1.173 |
| cifar100 | shufflenetv2_x1_0 | 1.181 | 3.386 | 1.726 | **1.070** | 1.074 | 1.073 |
| cifar100 | shufflenetv2_x1_5 | 1.073 | 3.371 | 1.519 | **1.022** | 1.026 | 1.024 |
| cifar100 | shufflenetv2_x2_0 | 0.998 | 2.976 | 1.513 | 0.972 | 0.980 | **0.972** |
| cifar100 | vgg11_bn | 1.518 | 3.444 | 1.693 | 1.248 | 1.256 | **1.248** |
| cifar100 | vgg13_bn | 1.333 | 3.061 | 1.825 | **1.112** | 1.123 | 1.116 |
| cifar100 | vgg16_bn | 1.640 | 2.998 | 1.536 | **1.113** | 1.120 | 1.113 |
| cifar100 | vgg19_bn | 1.798 | 2.927 | 1.530 | 1.138 | 1.137 | **1.133** |
| imagenet | mobilenetv2_120d | 0.956 | 3.834 | 1.824 | 0.903 | **0.897** | 0.921 |
| imagenet | repvgg_b3 | 0.835 | 3.476 | 1.760 | 0.828 | **0.814** | 0.840 |
| imagenet | tf_efficientnet_b8 | 0.665 | 2.548 | 1.447 | **0.582** | 0.587 | 0.653 |
| imagenet | vgg19_bn | 1.042 | 4.376 | 2.066 | 1.025 | **1.016** | 1.031 |