П

УДК 004.932.72 **DOI** 10.34757/2413-7383.2023.30.3.004

Я. С. Пикалёв[1], Т. В. Ермоленко[2]
[1]Федеральное государственное бюджетное научное учреждение
«Институт проблем искусственного интеллекта», г. Донецк
283048, г. Донецк, ул. Артема, 118-б
[2]Федеральное государственное бюджетное образовательное учреждение высшего образования «Донецкий государственный университет», г. Донецк
283001, г. Донецк, ул. Университетская, 24

О НЕЙРОННЫХ АРХИТЕКТУРАХ ИЗВЛЕЧЕНИЯ ПРИЗНАКОВ ДЛЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБЪЕКТОВ НА УСТРОЙСТВАХ С ОГРАНИЧЕННОЙ ВЫЧИСЛИТЕЛЬНОЙ МОЩНОСТЬЮ

Ya. S. Pikalyov[1], T. V. Yermolenko[2]
[1]Federal State Scientific Institution «Institute of Problems of Artificial intelligence», c. Donetsk
283048, Donetsk, Artema str., 118-b
[2]Federal State Educational Institution of Higher Education «Donetsk State University»
283001, Donetsk, University st, 24

ABOUT NEURAL ARCHITECTURES OF FEATURE EXTRACTION FOR THE PROBLEM OF OBJECT RECOGNITION ON DEVICES WITH LIMITED COMPUTING POWER

Данная работа посвящена исследованию эффективности различных моделей нейронных сетей в задачах обнаружения объектов и их классификации на устройствах с ограниченной вычислительной мощностью. Авторы используют двухэтапный подход на базе архитектуры Faster R-CNN для обнаружения объекта на изображении и его распознавания. Основным блоком в структуре Faster R-CNN, влияющим на качество и производительность всей системы, является базовая сеть. В работе представлены результаты численных исследований эффективности различных сетевых архитектур по таким критериям как разделяющая способность высокоуровневых признаков, точность классификации, количество занимаемой оперативной памяти, вычислительная сложность. Предложена интегральная оценка эффективности моделей, учитывающая указанные выше критерии. Наилучшее значение по интегральному критерию показала гибридная сеть EdgeNeXt-S, что свидетельствует о хорошем балансе этой модели между производительностью, робастностью и точностью в системах компьютерного зрения.

Ключевые слова: компьютерное зрение, обнаружение объектов, базовые сети, глубокое обучение, кластеризация, устройства с ограниченной вычислительной мощностью

This work is devoted to the study of the effectiveness of various neural network models in the tasks of object detection and classification on devices with limited computing power. The authors use a two-step approach based on the Faster R-CNN architecture to detect an object in an image and recognize it. The basic network is the main block in the Faster R-CNN structure that affects the quality and performance of the entire system. The paper presents the results of numerical studies of the effectiveness of various network architectures according to criteria such as the separating ability of high-level features, classification accuracy, the amount of RAM occupied, computational complexity. An integral assessment of the effectiveness of the models is proposed, taking into account the above criteria. The best value according to the integral criterion was shown by the hybrid network EdgeNeXt-S, which indicates a good balance of this model between performance, robustness and accuracy in computer systems

Key words: computer vision, object detection, backbone networks, deep learning, clusterisation, edge devices

Введение

С развитием методов глубокого обучения, становится возможной обработка крупномасштабных наборов данных. Кроме того, автоматическое обучение на большом наборе приводит к улучшению метрик, используя большое количество признаков. Для задач компьютерного зрения признаки извлекаются с использованием различных базовых сетей (*backbone network, BN*). Базовая сеть ¬ это общепризнанная архитектура или сеть, используемая для извлечения признаков и предварительно обученная на одной или нескольких задачах, как правило на большом наборе данных. Для задач компьютерного зрения выбор подходящей BN для извлечения признаков сложен из-за того, что в одних задачах используются определенные BN, а при использовании другой BN модель может показывать плохие результаты [1]. Учитывая вышесказанное, стоит отметить растущую потребность в использовании приложений компьютерного зрения, в том числе связанных с обнаружением объектов на устройствах с ограниченной вычислительной мощностью [2].

Постановка задачи исследования

Процесс извлечения информативных признаков, D, из BN можно описать следующим образом.

$$D = [F_D(x_i)], x \in R^{3\times H\times W}, i \in 0 \dots T \quad (1)$$

$$Y = [F_Y(D)], D \in R^{1\times d} \quad (2)$$

$$y = C[argmax(Y)], \quad (3)$$

где F_D – функция преобразования признаков x из размерности $3 \times H \times W$ в размерность $1 \times d$; F_Y – функция классификации; Y – вероятностное распределение для N-классов, T – размер пакета (батча) данных.

Цель настоящего исследования состоит в сравнении показателей производительности и качества метрик, используя различные BN, для дальнейшего применения наилучшей BN в устройствах с ограниченной вычислительной мощностью.

В работе ставятся следующие задачи:

1. Сформировать критерии и, используя их, отобрать BN.
2. Дать краткое описание отобранным BN.
3. Сформировать набор данных для дальнейших экспериментов.
4. Сформировать методику оценивания BN.
5. Выбрать BN на основе предложенной методики.

Обоснование выбора архитектур глубоких сетей для задачи распознавания изображений на устройствах с ограниченными ресурсами

Выбор рассматриваемых в работе архитектур сделан из тех соображений, что они обладают следующими качествами:

– указано точное количество параметров, при этом оно для выбранных архитектур не более 25 млн;
– качество распознавания, превышающее 75% по метрике top-1 на ImageNet [3]. Точность top-N означает, что в список из N наиболее вероятных попал правильный.

Для сравнения эффективности вышеуказанных архитектур использовались такие показатели как точность, GFLOPs (число операций с плавающей запятой, *floating-point operations*; GFLOPs = 10^9 FLOPs), количество параметров сети, которые сведены в табл. 1.

Таблица 1

№	Модель	Точность, %	Кол-во параметров, млн	GFLOPS
1	ConvNeXt-N [4]	81.9	15.6	2.45
2	DaViT-T [5]	82.8	28.3	4.5
3	EdgeFormer-S [6]	78.63	5	3.48
4	EdgeNeXt-S [7]	79.4	5.6	2.6
5	EfficientFormer-L [8]	83.5	26.1	7.6
6	EfficientNetV2-S [9]	83.9	24	8.8
7	MobileViTv3-S [10]	79.3	5.8	1.8
8	NextViT-S [11]	82.5	31.7	5.8
9	ResNet-50 [12]	75.3	25	3.8
10	TinyViT-21M [13]	84.8	21	4.3

Для краткости дальнейшего изложения название модели, указанное в табл. 1, будет заменяться соответствующим номером (например, BN1).

Формирование набора данных для обучения и тестирования моделей

Существующие наборы данных для оценки качества классификации изображений, в особенности как подзадачи обнаружения объектов из изображений, являются неоптимальными. Ключевыми факторами для этого являются: количество классов; общее количество изображений; наличие объектов различных масштабов. Дополнительно для упрощения данной подзадачи следует использовать классификацию одного объекта, т.к. в системах обнаружения объектов для классификации подаются отдельные кадры, предполагающие наличие единого объекта. Для этого в ходе данной работы был сформирован набор данных ObjectDet, основанный на наборе данных ObjectNet [14]. Для части набора данных ObjectNet в работе [15] были сформированы метаданные с информацией об ограничивающих рамках объекта. Используя эти метаданные, исходное изображение изменено при помощи операции вырезания (crop) объекта по ограничивающей рамке, с дальнейшим изменением его входного размера до значений 224×224. Для того чтобы сохранить исходные пропорции изображения недостающие пиксели были заполнены черным цветом (padding). Пример изменения изображения из ObjectNet (а) в изображение набора ObjectDet (б) приведен на рис. 3.

Дополнительно к обучающим данным ObjectDet применялась вероятностная аугментация (искусственное расширение набора данных). Использовались следующие техники аугментации для набора данных с изображениями:

1) горизонтальная инверсия с вероятностью (p) 0.5.
2) вертикальная инверсия, p=0.2;
3) выбор случайной прямоугольную области в изображении и удаление пикселей внутри [16], p=0.2;
4) случайное преобразование перспективы изображения, p=0.2

5) аффинные преобразования, p=0.5
6) сдвиг значений для каждого канала входного изображения, p=0.5
7) сдвиг значений яркости, контрастности, гаммы, p=0.5.
8) использование гауссова фильтра со случайными ядерными размерами, p=0.2

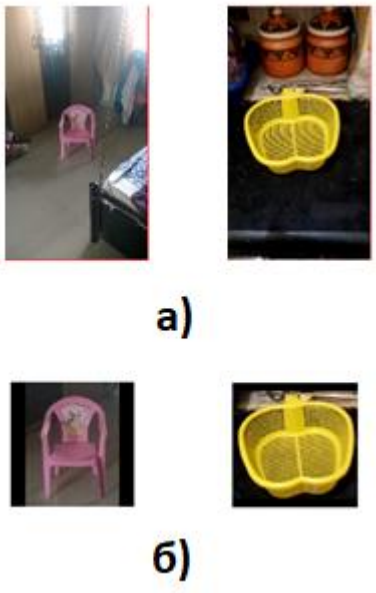

Рисунок 3 Пример формирования набора данных *ObjectDet*

Использование подобной техники аугментации при обучении на большом количестве эпох позволяет добиться высокой робастности и частично решает проблему переобучения.

Итоговый набор данных (табл. 2) разделен на обучающий (train), тестовый (test) и проверочный (val) в пропорциях 0.8-0.15-0.05. При разделении набора данных использовался подход стратификации (равномерного распределения классов) Общее количество классов составляет 172.

Таблица 2 – Общая характеристика набора данных ObjectDet

	Количество изображений	Общий размер, Гб
train	23415	1.35
test	4392	0.259
val	1725	0.102

Авторами предлагается методика оценивания BN, состоящая из следующих этапов.

1. Исследование разделяющей способности высокоуровневых признаков используемых архитектур. Для этого предлагается провести эксперимент по разбиению набора данных CIFAR-10 [17] на непересекающиеся подмножества (кластеры), используя отобранные BN.
2. Определения качества классификации изображений. Для этого предлагается провести эксперимент по обучению архитектуры классификации изображений на наборе данных ObjectDet, используя в качестве признаков выход из BN.
3. Определение занимаемой оперативной памяти для графического процессора.
4. Оценивание систем на основе критериев, полученных из предыдущих этапов.

Численные исследования эффективности моделей

Для исследования разделяющей способности высокоуровневых признаков, извлекаемых из рассматриваемых BN, был проведен эксперимент с применением визуализации при помощи модели UMAP [18], с использованием метод кластеризации высокоуровневых признаков при помощи метода K-средних [19]. В качестве критерия для оценки качества кластеризации в данной работе используется индекс Дэвиса-Булдина (Davies-Bouldin Index, DBI) [20], который вычисляется следующим образом:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j} R_{ij} \tag{4}$$

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}, \tag{5}$$

где s_i – среднее расстояние между каждой точкой кластера i и центроидом этого кластера; d_{ij} – расстояние между центроидами кластера i и j.

Результаты анализа эффективности BN по критериям DBI приведены в табл. 3.

Таблица 3 – Результаты оценки BN по критерию DBI

Модель	DBI
ConvNeXt-N	3.74
DaViT-T	3.79
EdgeFormer-S	23.82
EdgeNeXt-S	3.41
EfficientFormer-L	3.31
EfficientNetV2-S	3.23
MobileViTv3-S	1.77
NextViT-S	3.47
ResNet-50	4.29
TinyViT-21M	6.29

Модель для классификации изображений из набора ObjectDet идентична для всех BN, единственным отличным блоком является блок извлечения информативных признаков при помощи BN. На входной слой поступает вектор изображения размером 3×224×224. Для данного вектора, используя BN, извлекаются информативные признаки, D. После используется метод регуляризации (dropout) для решения проблемы переобучения сети. В качестве следующего слоя используется скрытый слой размерности d×N. Последним слоем является выходной слой, который выводит вероятностное распределение для каждого из N-классов.

В качестве оптимизатора для процесса обучения использовался подход Lion [21]; к которому дополнительно применяется механизм оптимизатора предвидения [22].

Общее количество эпох составляет 50. Размер пакета в данной работе составляет 32. В качестве основных метрик использовались перекрестная энтропия со сглаживанием классов, а также top-1 и top-5 точность.

Поскольку в данной работе исследуется возможность применимости BN в устройствах с ограниченной вычислительной мощностью, то одним из ключевых параметров является количество потребляемой памяти. Для этого был проведен эксперимент в режиме тестирования моделей на разном наборе батчей с целью определения занимаемой оперативной памяти для графического процессора (Video Random Access Memory, VRAM). Результаты отображены в табл. 4.

Таблица 4 – Количество VRAM, потребляемой моделями при размере пакета 64

Модель	VRAM, Гб
ConvNeXt-N	6.45
DaViT-T	7.10
EdgeFormer-S	8.89
EdgeNeXt-S	7.20
EfficientFormer-L	13.17
EfficientNetV2-S	8.36
MobileViTv3-S	8.96
NextViT-S	6.77
ResNet-50	5.36
TinyViT-21M	9.95

Результаты обучения и тестирования моделей отображены в табл. 5, где loss – значение функции потерь; loss_test – тестирования; acc1 – точность top-1; acc5 – точность top-5, train – режим обучения, test – режим тестирования.

Таблица 5 – Результаты обучения и тестирования BB

	loss		acc1		acc5	
	train	test	train	test	train	test
BN1	2.41	2.77	61.888	74.68	93.76	84.57
BN2	2.40	2.61	67.278	75.01	94.03	88.76
BN3	3.57	3.33	44.338	37.90	63.93	70.69
BN4	2.42	2.49	70.267	73.89	93.02	90.57
BN5	3.27	6.53	54.71	45.95	72.89	78.84
BN6	3.08	3.07	51.607	51.59	77.62	76.81
BN7	3.03	3.01	54.347	53.79	79.51	79.09
BN8	2.54	2.54	69.451	71.37	91.17	90.37
BN9	2.79	2.95	56.612	62.86	86.70	80.97
BN10	2.78	2.52	70.538	62.2	85.14	89.94

Дальнейшая оценка BN проводится на основе критериев, полученных из предыдущих этапов (табл. 1, 3-5):

1) acc1 на ImageNet (по максимальному значению);
2) GFLOPs (по минимальному значению);
3) кол-во параметров (params, по минимальному значению);
4) DBI (по максимальному значению);
5) количество VRAM при размере пакета 64 (по минимальному значению);
6) loss на тестовом наборе ObjectDet (по минимальному значению);
7) acc1 на тестовом наборе ObjectDet (по максимальному значению);
8) acc5 на тестовом наборе ObjectDet (по максимальному значению).

Для оценивания по вышеуказанным критериям использовалась 10-балльная шкала, в соответствии с количеством сравниваемых систем. Для этого на основе указанных выше показателей для каждой BN по отдельным критериям выставляется интегральная оценка S (максимальный балл выставляется для наилучшего показателя, минимальный – для наихудшего). Наилучшим вариантом считается тот, чья сумма набранных балов является наибольшей:

$$V = BN_k, k = 1 \cdots n, \quad (6)$$

$$k = argmax(S_i), i = 1 \cdots n, \quad (7)$$

$$S_i = \sum_{j=1}^{m} s_{1j}, \quad (8)$$

где S_i – одномерная матрица оценок $1 \times n$, n – общее количество BN, m – общее количество критериев.

Таблица 6 – Результаты оценивания BN по интегральному критерию

Модель	S
ConvNeXt-N	50
DaViT-T	42
EdgeFormer-S	27
EdgeNeXt-S	62
EfficientFormer-L	30
EfficientNetV2-S	32
MobileViTv3-S	44
NextViT-S	57
ResNet-50	39
TinyViT-21M	57

Выводы

Для обучения и тестирования моделей сформирован датасет ObjectDet, содержащий 172 класса и основанный на наборе данных ObjectNet с применением вероятностной аугментации, использование которой при обучении на большом количестве эпох позволяет добиться высокой робастности и частично решить проблему переобучения.

Проведена оценка разделяющей способности высокоуровневых признаков исследуемых архитектур с помощью индекса Дэвиса-Булдина, наихудшими показателями обладают гибридная модель EdgeFormer-S и трансформероподобная TinyViT-21M.

Помимо этого, для сравнения эффективности исследуемых архитектур использовались такие критерии как точность классификации, количество занимаемой оперативной памяти и интегральный критерий, учитывающий точность, вычислительную сложность, а также количество параметров модели.

Численные исследования тестирования моделей на разном наборе батчей для определения занимаемой оперативной памяти показали наилучшую эффективность архитектуры ConvNeXt-N, принадлежащей классу гибридных моделей, наиболее точной оказалась трансформероподобная модель DaViT-T. Наилучшее значение по интегральному критерию показала гибридная сеть EdgeNeXt-S, что свидетельствует о хорошем балансе данной модели между производительностью, робастностью и точностью в системах компьютерного зрения.

Предложенная методика интегрального оценивания нейросетевых моделей может в дальнейшем применяться исследователями для любого набора BN с целью выбора подходящей архитектуры для реализации в устройствах с ограниченной вычислительной мощностью.

Архитектуру EdgeNeXt-S, которая является наилучшим вариантом среди исследуемых моделей, рекомендуется применять разработчикам для систем компьютерного зрения, базирующихся на устройствах с ограниченной вычислительной сложностью.

Список литературы

1. Способ обучения нейронной сети управдения роботом / В М Зуев, О.А. Бутов, С.Б. Иванова, А.А. Никитина, С.И. Уланов. *Проблемы искусственного интеллекта*. 2021. Т. 2. № 21. С. 22-33.
2. Покинтелица А.Е. Проблемы и специфика редукции данных в автономных робототехнических системах. *Проблемы искусственного интеллекта*. 2023. Т. 1. № 28. С. 31-41.
3. Russakovsky O. ImageNet Large Scale Visual Recognition Challenge / O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei. *International Journal of Computer Vision*. 2015. Т. 115. № 3.
4. *ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoder*s / S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I.S. Kweon, S. Xie. 2023.
5. DaViT: Dual Attention Vision Transformers / M. Ding, B. Xiao, N. Codella, P. Luo, J. Wang, L. Yuan. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2022. Т. 13684 LNCS.
6. Zhang H., Hu W., Wang X. ParC-Net: Position Aware Circular Convolution with Merits from ConvNets and Transformer. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2022. Т. 13686 LNCS.
7. *EdgeNeXt: Efficiently Amalgamated CNN-Transformer Architecture for Mobile Vision Applications* / M. Maaz, A. Shaker, H. Cholakkal, S. Khan, S.W. Zamir, R.M. Anwer, F. Shahbaz Khan. 2023.
8. E*fficientFormer: Vision Transformers at MobileNet Speed* / Y. Li, G. Yuan, Y. Wen, J. Hu, G. Evangelidis, S. Tulyakov, Y. Wang, J. Ren. 2022.
9. Tan M., Le Q. V. *EfficientNetV2: Smaller Models and Faster Training*. 2021.
10. Wadekar S.N. MobileViTv3: Mobile-Friendly Vision Transformer with Simple and Effective Fusion of Local, Global and Input Features / S.N. Wadekar, A. Chaurasia. – 2022.
11. Next-ViT: Next Generation Vision Transformer for Efficient Deployment in Realistic Industrial Scenarios / J. Li, X. Xia, W. Li, H. Li, X. Wang, X. Xiao, R. Wang, M. Zheng, X. Pan. 2022.
12. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognitio*n. 2016. Тт. 2016-December.
13. TinyViT: Fast Pretraining Distillation for Small Vision Transformers / K. Wu, J. Zhang, H. Peng, M. Liu, B. Xiao, J. Fu, L. Yuan. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2022. Т. 13681 LNCS.
14. ObjectNet: A large-scale bias-controlled dataset for pushing the limits of object recognition models / A. Barbu, D. Mayo, J. Alverio, W. Luo, C. Wang, D. Gutfreund, J. Tenenbaum, B. Katz. *Advances in Neural Information Processing Systems*. 2019. Т. 32.
15. Borji A. *ObjectNet Dataset: Reanalysis and Correction*. 2020.
16. Random Erasing Data Augmentation / Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang. 2017.
17. Krizhevsky A. Learning Multiple Layers of Features from Tiny Images. *Science Department, University of Toronto, Tech*. 2009.
18. UMAP: Uniform Manifold Approximation and Projection / L. McInnes, J. Healy, N. Saul, L. Großberger. *Journal of Open Source Software*. 2018. Т. 3. № 29.
19. Coates A., Ng A.Y. Learning feature representations with K-means. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2012. Т. 7700 LECTURE NO.
20. Davies D.L., Bouldin D.W. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1979. Т. PAMI-1. № 2.
21. Symbolic Discovery of Optimization Algorithms / X. Chen, C. Liang, D. Huang, E. Real, K. Wang, Y. Liu, H. Pham, X. Dong, T. Luong, C.-J. Hsieh, Y. Lu, Q. V. Le. 2023.
22. Lookahead Optimizer: k steps forward, 1 step back / M. Zhang, J. Lucas, J. Ba, G.E. Hinton. *Advances in Neural Information Processing Systems*. 2019. С. 9593-9604.

References

1. Zuev V. M. Method for learning neural network for robot control/ V. M. Zuev, O.A. Butov, S.B.Ivanova, A.A.Nikitina, S.I. Ulanov. *Problems of Artificial Intelligence*. 2021. Vol. 2. № 21. P. 22-33.
2. Pokintelitsa A.E. Problems and features of data reduction in autonomous robotic systems / A.E.Pokintilitsa. P*roblems of Artificial Intelligence*. 2023. Vol. 1. № 28. P. 31-41.

3. Russakovsky O. ImageNet Large Scale Visual Recognition Challenge / O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei. I*nternational Journal of Computer Vision*. 2015. Vol. 115. № 3.
4. Woo S. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders / S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I.S. Kweon, S. Xie. 2023.
5. Ding M. DaViT: Dual Attention Vision Transformers / M. Ding, B. Xiao, N. Codella, P. Luo, J. Wang, L. Yuan. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2022. Vol. 13684 LNCS.
6. Zhang H. ParC-Net: Position Aware Circular Convolution with Merits from ConvNets and Transformer / H. Zhang, W. Hu, X. Wang. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2022. Vol. 13686 LNCS.
7. *Maaz M. EdgeNeXt: Efficiently Amalgamated CNN-Transformer Architecture for Mobile Vision Applications* / M. Maaz, A. Shaker, H. Cholakkal, S. Khan, S.W. Zamir, R.M. Anwer, F. Shahbaz Khan. 2023.
8. Li Y. E*fficientFormer: Vision Transformers at MobileNet Speed* / Y. Li, G. Yuan, Y. Wen, J. Hu, G. Evangelidis, S. Tulyakov, Y. Wang, J. Ren. 2022.
9. Tan M. *EfficientNetV2: Smaller Models and Faster Training* / M. Tan, Q. V. Le. 2021.
10. Wadekar S.N. *MobileViTv3: Mobile-Friendly Vision Transformer with Simple and Effective Fusion of Local, Global and Input Features* / S.N. Wadekar, A. Chaurasia. 2022.
11. Li J. *Next-ViT: Next Generation Vision Transformer for Efficient Deployment in Realistic Industrial Scenarios* / J. Li, X. Xia, W. Li, H. Li, X. Wang, X. Xiao, R. Wang, M. Zheng, X. Pan. 2022.
12. He K. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2016. Vol. 2016-December.
13. Wu K. TinyViT: Fast Pretraining Distillation for Small Vision Transformers / K. Wu, J. Zhang, H. Peng, M. Liu, B. Xiao, J. Fu, L. Yuan. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2022. Vol. 13681 LNCS.
14. Barbu A. ObjectNet: A large-scale bias-controlled dataset for pushing the limits of object recognition models / A. Barbu, D. Mayo, J. Alverio, W. Luo, C. Wang, D. Gutfreund, J. Tenenbaum, B. Katz. *Advances in Neural Information Processing Systems*. 2019. Vol. 32.
15. Borji A. *ObjectNet Dataset: Reanalysis and Correction* / A. Borji. 2020.
16. Zhong Z. *Random Erasing Data Augmentation* / Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang. 2017.
17. Krizhevsky A. Learning Multiple Layers of Features from Tiny Images / A. Krizhevsky. *Science Department, University of Toronto, Tech.* 2009.
18. McInnes L. UMAP: Uniform Manifold Approximation and Projection / L. McInnes, J. Healy, N. Saul, L. Großberger. *Journal of Open Source Software*. 2018. Vol. 3. № 29.
19. Coates A. Learning feature representations with K-means / A. Coates, A.Y. Ng. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2012. Vol. 7700 LECTURE NO.
20. Davies D.L. A Cluster Separation Measure / D.L. Davies, D.W. Bouldin. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1979. Vol. PAMI-1. № 2.
21. Chen X. *Symbolic Discovery of Optimization Algorithms* / X. Chen, C. Liang, D. Huang, E. Real, K. Wang, Y. Liu, H. Pham, X. Dong, T. Luong, C.-J. Hsieh, Y. Lu, Q. V. Le. 2023.
22. Zhang M. Lookahead Optimizer: k steps forward, 1 step back / M. Zhang, J. Lucas, J. Ba, G.E. Hinton. *Advances in Neural Information Processing Systems*. 2019. P. 9593-9604.

RESUME

Ya. S. Pikalyov, T. V. Yermolenko

About Neural Architectures of Feature Extraction For The Problem Of Object Recognition On Devices With Limited Computing Power

A number of papers were devoted to the study of neural architectures of feature extraction (basic network) for the object recognition task, in which their application for a number of computer vision tasks, such as image classification, object detection, face

recognition, panoptic segmentation, action recognition, etc., was discussed. Nevertheless, these papers do not disclose the issue of the use of basic networks on devices with limited computing power. It is worth noting that the methodology for evaluating basic networks also remains an open task.

To compare the efficiency of the above architectures, such indicators as accuracy, the number of floating-point operations, and the number of network parameters were used. To study the separating ability of high-level features extracted from the underlying networks under consideration, an experiment was conducted on the CIFAR-10 dataset using visualization using the UMAP model, using the clustering method. high-level features using the K-means method. The Davies-Bouldin index is used as a criterion for assessing the quality of clustering in this paper. To evaluate the underlying networks for the classification task on the ObjectDet dataset, metrics such as the value of the loss function and the accuracy indicator (the ratio of correctly predicted classes to all) were used. As a final metric for the evaluation of basic networks, an integral assessment (on a 10-point scale) was used based on the above criteria. The best option is the one whose sum of points scored is the largest.

Numerical studies of testing models on a different set of patches to determine the occupied RAM showed the best efficiency of the ConvNeXt-N architecture belonging to the class of hybrid models, the most accurate was the transformer-like DaViT-T model. The Edge NeXt-S hybrid network showed the best value according to the integral criterion, which indicates a good balance of this model between performance, robustness and accuracy in computer vision systems.

The authors selected a number of modern basic networks applicable in devices with limited computing capacity.The results of numerical studies of the effectiveness of various network architectures according to criteria such as the separating ability of high-level features, classification accuracy, the amount of RAM occupied, computational complexity are presented. A method of integral evaluation of neural network models is proposed, taking into account these indicators, which can be further applied by researchers for any set of basic networks in order to select a suitable architecture for implementation in devices with limited computing power.

РЕЗЮМЕ

Я. С. Пикалёв, Т. В. Ермоленко

О нейронных архитектурах извлечения признаков для задачи распознавания объектов на устройствах с ограниченной вычислительной мощностью

Вопросам исследования нейронных архитектур извлечения признаков (базовая сеть) для задачи распознавания объектов был посвящен ряд работ, в которых обсуждалось их применение для ряда задач компьютерного зрения, таких как классификация изображений, обнаружение объектов, распознавание лиц, паноптическая сегментация, распознавание действий и т.д. Тем не менее в этих работах не раскрыт вопрос применения базовых сетей на устройствах с ограниченной вычислительной мощностью. Стоит отметить, что методика оценки базовых сетей также остаётся открытой задачей.

Для сравнения эффективности вышеуказанных архитектур использовались такие показатели как точность, число операций с плавающей запятой, количество параметров сети. Для исследования разделяющей способности высокоуровневых

признаков, извлекаемых из рассматриваемых базовых сетей, был проведен эксперимент на наборе данных CIFAR-10 с применением визуализации при помощи модели UMAP, с использованием метода кластеризации. высокоуровневых признаков при помощи метода K-средних. В качестве критерия для оценки качества кластеризации в данной работе используется индекс Дэвиса-Булдина. Для оценки базовых сетей для задачи классификации на наборе данных ObjectDet использовались такие метрики как значение функции потерь и показатель точности (отношение правильно предсказанных классов ко всем). В качестве итоговой метрики для оценивания базовых сетей использовалась интегральная оценка (по 10-бальной шкале) на основе вышеуказанных критериев. Наилучшим вариантом считается тот, чья сумма набранных балов является наибольшей.

Численные исследования тестирования моделей на разном наборе батчей для определения занимаемой оперативной памяти показали наилучшую эффективность архитектуры ConvNeXt-N, принадлежащей классу гибридных моделей, наиболее точной оказалась трансформероподобная модель DaViT-T. Наилучшее значение по интегральному критерию показала гибридная сеть EdgeNeXt-S, что свидетельствует о хорошем балансе данной модели между производительностью, робастностью и точностью в системах компьютерного зрения.

Авторами был отобран ряд современных базовых сетей, применимых в устройствах с ограниченной вычислительной способностью. Представлены результаты численных исследований эффективности различных сетевых архитектур по таким критериям как разделяющая способность высокоуровневых признаков, точность классификации, количество занимаемой оперативной памяти, вычислительная сложность. Предложена методика интегрального оценивания нейросетевых моделей, учитывающая эти показатели, которая может в дальнейшем применяться исследователями для любого набора базовых сетей с целью выбора подходящей архитектуры для реализации в устройствах с ограниченной вычислительной мощностью.

Пикалёв Ярослав Сергеевич – кандидат технических наук, научный сотрудник отдела Интеллектуальных Робототехнических Систем, Федерального государственного бюджетного научного учреждения "Институт проблем искусственного интеллекта". *Область научных интересов:* Цифровая обработка сигналов, анализ данных, распознавание образов, обработка естественного языка, компьютерное зрение, машинное обучение, нейронные сети

Ермоленко Татьяна Владимировна – кандидат технических наук, доцент кафедры компьютерных технологий физико-технического факультета, Федерального государственного бюджетного образовательного учреждения высшего образования "Донецкого Национального Университета". *Область научных интересов:* Цифровая обработка сигналов, анализ данных, дискретная математика, теория алгоритмов, распознавание образов, обработка естественного языка, компьютерное зрение, машинное обучение, нейронные сети

Статья поступила в редакцию 19.04.2023.