И. И. ЖУРАВЛЕВ, О. А. МИЛОСЕРДОВ, А. В. МАКАРЕНКО
Институт проблем управления им. В.А.Трапезникова РАН, Москва

ПРЕДСТАВЛЕНИЕ ИЗОБРАЖЕНИЙ В ПРОСТРАНСТВЕ «ФОРМА–ТЕКСТУРА» ДЛЯ РЕШЕНИЯ НЕКОТОРЫХ ЗАДАЧ МАШИННОГО ЗРЕНИЯ МЕТОДАМИ ГЛУБОКОГО ОБУЧЕНИЯ

В настоящей работе проведён анализ существующих подходов к обучению нейросетей для решения задач машинного зрения. Установлено, что эти подходы имеют ряд недостатков, которые существенно затрудняют интерпретацию и управление признаками изображений в скрытом пространстве нейросетей. С целью исправления ситуации предложен оригинальный подход, в основе которого лежит перевод изображения в признаковое описание в пространстве ФОРМАхТЕКСТУРА. При этом показано, что данное пространство является в определённом роде базовым при решении широкого спектра задач в системах компьютерного зрения.

Введение. В настоящий момент области и задачи применения систем компьютерного зрения испытывают взрывной рост [1]. При этом неуклонно усиливаются требования к качеству и устойчивости функционирования алгоритмов обработки информации в этих системах, на фоне всё более сложных условий их применения [2]. Основным подходом, в части построения алгоритмов обработки информации в системах технического зрения, на текущее время являются предварительно обучаемые модели на основе глубоких искусственных нейронных сетей (ИНС) [3].

Построение контуров обработки целевой информации в системах технического зрения на основе глубоких ИНС имеет под собой ряд как положительных, так и отрицательных моментов. Причём недостатки подобных алгоритмов, в определённой мере, являются продолжением их достоинств. Так, глубокие ИНС позволяют полностью инкапсулировать (автоматизировать) вопросы синтеза первичных и вторичных информативных признаков служащих для решения тех или иных задач анализа фото- и видеоизображений. В свою очередь, подобная инкапсуляция приводит к ситуации построения аналитического алгоритма в виде «чёрного ящика», что для некоторых задач является неприемлемым, в силу неинтерпретируемости решений. Кроме того, отсутствие доступа к информативным признакам не позволяет проводить тюнинг моделей и/или их комплексирование, что для ряда задач и приложений также является критическим недостатком.

В настоящей работе, с целью устранения вышеизложенных ограничений, предлагается оригинальный подход, в основе которого лежит перевод изображения в признаковое описание в пространстве ФОРМАхТЕКСТУРА (далее пространство SxT). При этом, как показано далее, данное пространство является в определённом роде базовым при решении широкого спектра задач в системах компьютерного зрения.

Недавние исследования. Для повышения качества моделей большой акцент исследователи делают на архитектурах ИНС и подходах к их обучению. В задачах классификации важнейшим этапом является извлечение признаков. Авторы статьи [4] предлагают подход к обучению, основанный на self-supervised learning. В статьях [5], [6], [7] повышают качество модели, используя специальную функцию потерь для формирования эмбеддингов изображений. На основе архитектуры трансформеров в статье [8] изображения представляют в виде последовательности патчей и обрабатывают их, используя механизмы внимания.

Однако все вышеперечисленные подходы имеют общий недостаток – невозможность интерпретации и управления признаками скрытого пространства. Таким образом, предложенный в данной работе метод закладывает фундамент для создания моделей, свободных от данных недостатков.

Форма и текстура. Введем, в первом приближении, определения формы и текстуры.

Определение 1. Форма объекта S – характеристики границы объекта, инвариантные, как минимум, относительно операторов смещения, поворота, масштаба и отражения.

Определение 2. Пусть задано множество объектов $E = \{e: e \subset O\}$. Назовем $e \in E$ элементом текстуры. Взаимное расположение элементов текстуры e есть рисунок текстуры. Тогда текстура объекта T есть рисунок этого объекта.

Метод. Как было сказано во введении, мы стремимся сделать модель обучения более интерпретируемой, разделив пространство признаков изображения на два независимых подпространства S и T. Данный метод предлагается реализовать следующим образом (см. рис. 1). Обучая ИНС сеть как многоклассовый классификатор предсказывать класс формы и текстуры объекта, используя бинарную кросс энтропию ($\mathcal{L}^*_{BCE}$), где * либо S, либо T, мы одновременно воздействуем на заранее определенные две области ИНС с помощью специальной функции потерь $\mathcal{L}_S$ и $\mathcal{L}_T$, заставляя таким образом формироваться в одной из них только признаки, описывающие форму, а в другой – только текстуру. Классификатор состоит из энкодера f, реализованного в виде свёрточной нейронной сети и классификатора c, реализованного в виде многослойного перцептрона. Выходы классификатора S_{out}, T_{out} пропускаются через сигмоиды. Архитектура f и c представлена на рис. 2. N – размерность пространства признаков, соотвественно M и $(N - M)$ – размерность подпространств формы и текстуры.

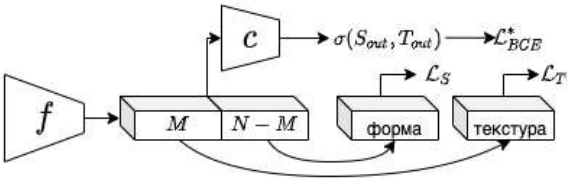

Рис. 1. Схема метода

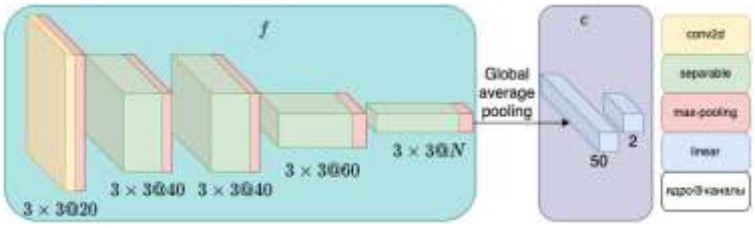

Рис. 2. Архитектура энкодера и классификатора

Датасет. Для первичного исследования предложенного метода использовался синтетический датасет (с целью повышения управляемости эксперимента) из заранее заданных классов формы $Y^S = \{\text{треугольник, квадрат, крест, месяц, эллипс}\}$ и текстуры $Y^T = \{\text{шум, гладкость, градиент, зебра, точки}\}$. Таким образом, имелось всего 25 классов различных объектов, представляющих собой геометрические фигуры с различной текстурой. При этом решалась задача бинарной классификации. Целевой класс представлял собой треугольник с текстурой зебра. Введем обозначения. Пусть x_i – объект класса, которому соответствуют метки классов $y_i^S \in Y^S$ и $y_i^T \in Y^T$. Если объект обладает целевой формой или текстурой, то $y_i^* = 1$, иначе $y_i^* = 0$. Изображения имеют разрешение 256×256 пикселей в черно-белом формате.

Метрика кластеризации. Используемая метрика кластеризации – силуэт [9], оценивающий разделённость кластеров. В экспериментах метрику разделимости кластеров в исходном пространстве обозначим S, разделимость кластеров целевой формы, но произвольной текстуры и наоборот обозначим S^*_{ST}. Отделимость кластеров целевой формы или текстуры от нецелевой обозначим S^*.

Функции потерь. Для формирования независимых подпространств использовалась center loss с градиентным обновлением центров кластеров [6]. Выбор данной функции потерь обусловлен тем, что ее оптимизация занимает малое время. Благодаря градиентному обновлению центров она учитывает информацию обо всём датасете и оценка центров получается минимально смещенной. Также данная функция потерь способна формировать

эмбеддинг изображений эллипсоидной формы, что позволяет использовать выбранную метрику кластеризации. Таким образом, оптимизировалась суммарная функция потерь:

$$\mathcal{L} = \alpha\mathcal{L}_{BCE}^{S} + \beta\mathcal{L}_{BCE}^{T} + \gamma\mathcal{L}_{S} + \theta\mathcal{L}_{T}, \quad (1)$$

где $\alpha, \beta, \gamma, \theta$ – коэффициенты вклада функции потерь.

Результаты экспериментов. Эксперименты производились с использованием метода оптимизации adam [10] с переменной скоростью обучения. На каждый эксперимент приходилось по 3 запуска, размерности исходного пространства были 10, 20, 40, 80, размерности подпространств равны, $\alpha = \beta = 1$. В данной работе приведены результаты для размерности пространства 10. Результаты приведены в таблице для первого запуска. Метрика $F1_*$ используется для оценки качества классификации формы и текстуры.

Таблица

Параметры обучения и значения метрик

N	M	γ		θ		$F1_S$	$F1_T$	S	S_{ST}^{S}	S^S	S_{ST}^{T}	S^T
		Эпоха										
		1-20	20-30	1-20	20-30							
10	5	0,01	0,1	0,01	0,1	1	1	0,84	0,06	0,84	0,02	0,91

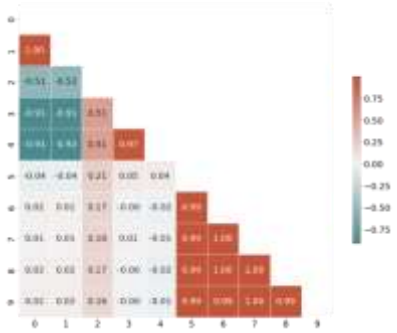

Рис. 3. Матрица корреляций. Признаки 0-4 описывают форму, 5-9 описывают текстуру.

Исследование независимости S и T. Для исследования линейной связи вычисляется матрица корреляций. Исходя из результатов можно сделать вывод, что линейная связь между признаками подпространств отсутствует (рис. 2), однако наблюдается «просачивание» признака формы в признаки текстуры, так как 2-й признак формы имеет большую относительно других величину коэффициента корреляции. Также проверим независимость S от T при помощи морфинга [11] целевой текстуры. Исходя из проекций на пространство 3-х главных компонент исходных пространств можно сделать вывод, что подпространства действительно независимы, поскольку при плавном изменении текстуры точка перемещается из кластера целевой текстуры в кластер нецелевой текстуры, при этом точка в подпространстве формы колеблется в пределах кластера целевой формы.

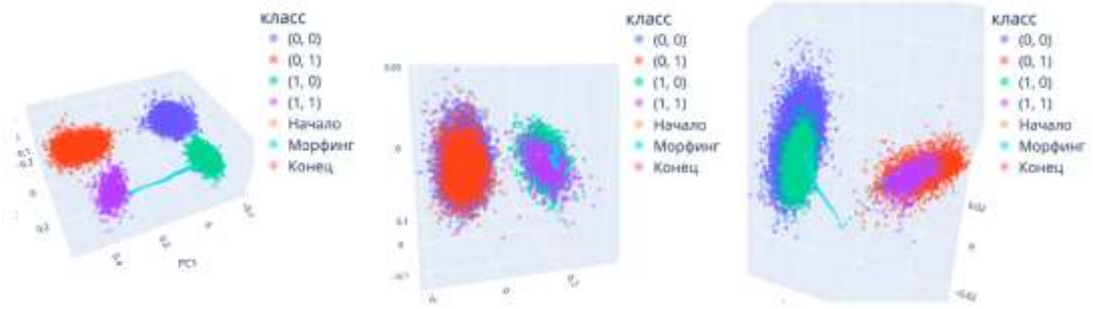

Рис. 4. Переход вектора признаков изображения во время морфинга текстуры в пространстве 3-х главных компонент

Заключение. В данной работе предложен алгоритм, позволяющий разделить пространство признаков на два независимых подпространства S и T. Данный алгоритм был исследован на синтетических данных. При помощи анализа коррелированности признаков и морфинга текстуры была продемонстрирована независимость подпространства формы от подпространства текстуры. Наше предположение о возможности разделения исходного пространства на два независимых подпространства S и T подтверждается экспериментами и позволяет продолжить исследования по созданию интерпретируемых и управляемых моделей

глубокого обучения для машинного зрения. В последующих работах мы предполагаем убедиться в независимости T от S при помощи морфинга формы, а также перенести результаты, полученные на синтетических данных, на реальные данные и задачи.

ЛИТЕРАТУРА

1. Shawahna A., Sait S., El-Maleh A. FPGA-Based Accelerators of Deep Learning Networks for Learning and Classification: *A Review. IEEE Access*. 2018. Pp. 1-1.
2. Jiao L., Zhao J. A Survey on the New Generation of Deep Learning in Image Processing. IEEE Access. 2019. Pp. 1-1.
3. Jena B., Nayak G., Saxena S. Convolutional neural network and its pretrained models for image classification and object detection: A survey. *Concurrency and Computation: Practice and Experience*. 2021. Vol. 34.
4. Gidaris S., Singh P., Komodakis N Unsupervised Representation Learning by Predicting Image Rotations. arXiv. 2018.
5. Schroff F., Kalenichenko D., Philbin J. FaceNet: A unified embedding for face recognition and clustering. *IEEE*. 2015.
6. Wen Y., Zhang K., Li Z., Qiao Y. A Discriminative Feature Learning Approach for Deep Face Recognition. Springer International Publishing. 2016. Pp. 499-515.
7. Sohn K., Li C.-L., Yoon J., Jin M., Pfister, T. Learning and Evaluating Representations for Deep One-class Classification. arXiv. 2020.
8. Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv. 2020.
9. Rousseeuw P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987. Vol. 20. Pp. 53-65.
10. Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization. arXiv. 2014.
11. URL: https://github.com/ddowd97/Python-Image-Morpher.

I.I.Zhuravlev, O.A.Miloserdov, A.V.Makarenko, (V.A. Trapeznikov (Institute of Control Sciences of Russian Academy of Sciences, Moscow)

Representation of images in the “SHAPExTEXTURE” space for solving some machine vision problems by deep learning methods

In this paper, the analysis of existing approaches to training neural networks for solving machine vision problems is carried out. It is established that these approaches have a few disadvantages that significantly complicate the interpretation and control of image features in the hidden space of neural networks. To correct the situation, an original approach is proposed, which is based on the translation of the image into a feature description in the SHAPExTEXTURE space. At the same time, it is shown that this space is in a certain way basic for solving a wide range of problems in computer vision systems.