

А.А. ДАШКЕВИЧ, канд. техн. наук, доц., доцент кафедры геометрического моделирования и компьютерной графики, НТУ "ХПИ"

ИССЛЕДОВАНИЕ МОДЕЛЕЙ СВЕРТОЧНЫХ АВТОЭНКODЕРОВ ДЛЯ ВЫДЕЛЕНИЯ ПРИЗНАКОВ В НАБОРАХ СТЕРЕОИЗОБРАЖЕНИЙ

В работе проведен процесс моделирования обучения без учителя сверточных автоэнкодеров для выделения признаков в наборах стереопар. Исследовано влияние количества обучаемых фильтров и топологии автоэнкодера на точность восстановления изображений. Определены конфигурации автоэнкодеров для точного восстановления входных изображений. Ил.: 2. Табл.: 1. Библиогр.: 13 назв.

Ключевые слова: обучение без учителя; сверточный автоэнкодер; выделение признаков; стереопара; восстановление изображений.

Постановка проблемы. В задачах восстановления трехмерных моделей сцены по наборам стереоизображений важным этапом является выделение ключевых точек на каждом изображении стереопары. Существующие алгоритмы обработки и выделения признаков требуют проводить предварительную обработку изображений. В то же время, в последние годы происходит стремительное развитие моделей искусственных нейронных сетей, в частности, сверточных нейронных сетей [1, 2], в которых процессы предварительной обработки и выделения признаков являются следствием их механизма работы. Однако, в работе таких сетей необходимо наличие пар "обучающий пример – желаемый выход сети", что в случае стереоизображений приводит к необходимости предварительного выделения карт глубины. Это требует дополнительного проведения процесса калибровки камеры с целью определения внешних и внутренних параметров камеры. Соответственно, если исключить данный процесс из технологической цепочки при восстановлении трехмерной сцены, то возможно существенное упрощение вычислений. Данный подход предлагается реализовать с помощью метода обучения без учителя, который требует только наличия обучающих примеров.

Анализ последних исследований. Одним из методов обучения без учителя является автоэнкодер [3], который позволяет эффективно находить зависимости во входных данных. В данной работе рассматриваются модели так называемых сверточных автоэнкодеров [4], которые в большей степени подходят для задач обработки изображений. Важным свойством этого класса нейронных сетей является то, что в

процессе обучения в скрытых слоях формируются сжатые представления входных данных [5, 6], что позволяет, например, проводить очистку изображений от шума [7] или проводить генерацию новых изображений [8]. В работах [9, 10] показано применение автоэнкодеров при восстановлении карт глубины по одиночным изображениям, приводятся варианты функции потерь для оценки качества восстановления. В работе [11] было проведено исследование влияния параметров топологии многослойного персептрона на его работу, были выявлены ключевые параметры, влияющие на сходимость обучения персептрона, предложенный подход может быть применен и для сверточных нейронных сетей. В работе [12] при исследовании сверточных нейронных сетей было установлено, что наращивание глубины сети, при относительно небольшом количестве используемых фильтров свертки, и совместно с использованием слоев субдискретизации приводит к улучшению качества классифицированных изображений при сниженной вычислительной нагрузке.

Цель работы. Изучение влияния параметров и топологии сверточного автоэнкодера на процесс выделения признаков при обучении без учителя.

Методология исследования. Исследовалась работа 5 различных топологий сверточных автоэнкодеров сетей для решения задачи восстановления входного изображения. Основные параметры разработанных моделей приведены в табл. 1. В качестве обучающего набора данных был использован набор "Middlebury Stereo 2014" [13], который содержит стереопары изображений, карты глубин, а также калибровочные параметры камеры (рис. 1). Так как данный набор содержит изображения достаточно больших размеров, то было проведено расширение набора путем вырезания из исходных изображений участков меньшего размера (128×128 пикселей) в случайных местах, таким образом из 23 исходных изображений набора были получены обучающая выборка из 1380 изображений и тестовая выборка из 920 изображений.

Для каждой модели проводилось 30 итераций обучения. Затем процесс обучения повторялся два раза, каждый раз с новой обучающей выборкой. После каждого из этапов проводилось тестирование на тестовой выборке. Оценка производилась как среднеквадратичное отклонение исходных данных от восстановленных. Результаты обучения для разработанных моделей приведены в табл. 1. На рис. 2 показаны примеры исходного и восстановленных изображений для всех моделей.

Таблица

Параметры разработанных моделей и точность восстановления входных изображений

Номер модели	Архитектура, размерность входов, карт признаков и выходов, число обучаемых параметров	Точность
Общая архитектура моделей 1-4 Conv-Pool-Conv-Pool-Conv- DeConv-UpSamp-DeConv-UpSamp-DeConv ¹		
1	(128;3) ² -(124;16)-(62;16)-(60;16)-(30;16)-(28;16)-(30;16)-(60;16)-(62;16)-(124;16)-(128;3), 11699	50.01
2	(128;3)-(124;16)-(62;16)-(60;16)-(30;16)-(28;1)-(30;16)-(60;16)-(62;16)-(124;16)-(128;3), 7364	674.58
3	(128;3)-(124;32)-(62;32)-(60;32)-(30;32)-(28;3)-(30;32)-(60;32)-(62;32)-(124;32)-(128;3), 25094	74.42
4	(128;3)-(124;8)-(62;8)-(60;8)-(30;8)-(28;3)-(30;8)-(60;8)-(62;8)-(124;8)-(128;3), 2822	179.28
Архитектура модели 5 Conv-Pool-Conv-Pool-Conv-Pool-Conv- DeConv-UpSamp-DeConv-UpSamp-DeConv-UpSamp-DeConv		
5	(128;3)-(124;16)-(62;16)-(60;16)-(30;16)-(28;16)-(14;16)-(12;3)-(14;16)-(28;16)-(30;16)-(60;16)-(62;16)-(124;16)-(128;3), 12582	820.74
¹ - Conv – сверточный блок; Pool – слой субдискретизации; DeConv – слой развертки; UpSamp – слой увеличения размерности в 2 раза; ² - Размерность – 3 карты признаков размером 128×128 пикселей.		



Рис. 1. Примеры изображения и соответствующей карты глубины

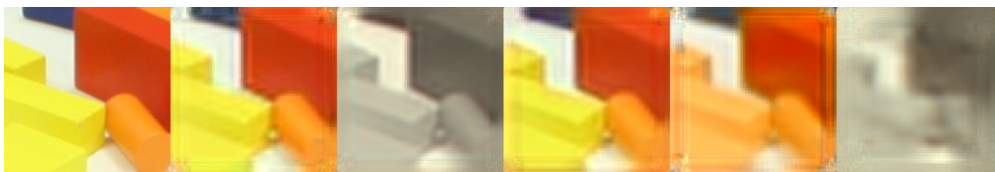


Рис. 2. Результаты восстановления. Слева-направо: входное изображение, модели 1-5

Выводы и перспективы дальнейших исследований. Из исследования можно увидеть, что:

- автоэнкодеры восстанавливают входное изображение путем сжимающего преобразования, за счет которого происходит выделение существенных признаков;

- рост числа фильтров в слоях свертки не приводит к существенному увеличению точности, но приводит к существенному росту количества вычислений (модель № 3), обратное также является верным (модель № 4);

- при низких размерностях карт признаков на границе кодирующей и декодирующей частей существенно ухудшается точность решения, что можно увидеть на примерах моделей № 2 и № 5, в которых, кроме снижения качества восстановленного изображения, теряется также и информация о цвете;

- автоэнкодер показывает эффективные результаты даже при небольшой глубине сети, что повышает скорость обработки изображений.

Это подтверждают результаты модели № 1, которая при относительно небольшом количестве обучаемых параметров продемонстрировала наилучшие результаты после 90 итераций обучения на тестовых данных. При этом модель № 3 демонстрирует практически такой же результат, как и модель № 1, но имеет при этом в два раза большее число обучаемых весов, что отражается на скорости ее вычислений.

В качестве важного направления дальнейших исследований представляется моделирование автоэнкодеров, которые будут обрабатывать два изображения стереопары за один проход и восстанавливать на выходах карту глубин сцены.

Список литературы: 1. *LeCun Y.* Gradien-based learning applied to document recognition / *Y. LeCun, L. Bottou, Y. Bengio, P. Haffner* // Proceedings of the IEEE. – 1998. – Vol. 86. – Issue 11. – P. 2278-2324. 2. *Krizhevsky A.* ImageNet Classification with Deep Convolutional Neural Networks / *A. Krizhevsky, I. Sutskever, G.E. Hinton*. // NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems. – 2012. – P. 1097-1105. 3. *Hinton G.* Reducing the Dimensionality of Data with Neural Networks / *G. Hinton, R. Salakhutdinov* // Science. – 2006. – Vol. 313. – № 5786. – P. 504-507. 4. *Masci J.*

Stacked convolutional auto-encoders for hierarchical feature extraction / *J. Masci, U. Meier, D. Cireşan, J. Schmidhuber* // In Proceedings of the 21th international conference on Artificial neural networks (ICANN'11). – 2011. – Vol. – Part I. – Springer-Verlag, Berlin, Heidelberg. – P. 52-59. **5.** *Adams R.P.* On Nonparametric Guidance for Learning Autoencoder Representations / *R.P. Adams, H. Larochelle, J. Snoek* // Proceedings of Machine Learning Research. – 2012. – Vol. 22: AISTATS. – P. 1073-1080. **6.** *Makhzani A.* Winner-Take-All Autoencoders / *A. Makhzani, B. Frey* // NIPS'15 Proceedings of the 28th International Conference on Neural Information Processing Systems. – 2015. – Vol. 2. – P. 2791-2799. **7.** *Stowell D.* Denoising without access to clean data using a partitioned autoencoder / *D. Stowell, R. Turner* // CoRR, abs/1509.05982. – 2015. **8.** *Brox T.* Generating Images with Perceptual Similarity Metrics based on Deep Networks / *T. Brox, A. Dosovitskiy* // NIPS'16 Proceedings of the 29th International Conference on Neural Information Processing Systems. – 2016. – P. 658-666. **9.** *Carneiro G.* Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue / *G. Carneiro, R. Garg, B.G. Kumar, I.D. Reid* // European Conference on Computer Vision 2016: Computer Vision – ECCV. – 2016. – P. 740-756. **10.** *Bernard F.* MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction / *F. Bernard, P. Garrido, H. Kim, P. Pérez, A. Tewari, C. Theobalt, M. Zollhöfer* // CoRR, abs/1703.10580. – 2017. **11.** *Дашкевич А.О.* Дослідження багатопарових нейронних мереж для автоматичного виділення ознак при вирішенні задачі розпізнавання образів / *А.О. Дашкевич* // Науковий вісник ТДАТУ. – Мелітополь: ТДАТУ. – 2016. – Вип. 6. – Т. 2. – С. 134-139. **12.** *Дашкевич А.А.* Исследование архитектуры сверточных нейронных сетей для решения задачи классификации изображений / *А.А. Дашкевич* // Системы обработки информации. – Харьков: ХНУПС ім. І. Кожедуба. – 2017. – Вип. 4 (150). – С. 109-112. **13.** *Scharstein D.* High-resolution stereo datasets with subpixel-accurate ground truth / *D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesić, X. Wang, P. Westling* // GCPR. – 2014. – Vol. 8753 of Lecture Notes in Computer Science. – P. 31-42.

References:

1. LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998), "Gradien-based learning applied to document recognition". *Proceedings of the IEEE*, Vol. 86, Issue 11, Nov. 1998, pp. 2278 - 2324.
2. Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012), "ImageNet Classification with Deep Convolutional Neural Networks". *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 1097-1105.
3. Hinton, G.E. and Salakhutdinov, R.R. (2006), "Reducing the Dimensionality of Data with Neural Networks", *Science*, 28 July 2006, Vol. 313, No. 5786, pp. 504-507.
4. Masci, J., Meier, U, Cireşan, D. and Schmidhuber, J. (2011), "Stacked convolutional auto-encoders for hierarchical feature extraction". *In Proceedings of the 21th international conference on Artificial neural networks (ICANN'11)*, Vol. Part I, Springer-Verlag, Berlin, Heidelberg, pp. 52-59.
5. Adams, R.P., Larochelle, H. and Snoek, J. (2012), "On Nonparametric Guidance for Learning Autoencoder Representations". *Proceedings of Machine Learning Research*, Vol. 22: AISTATS, pp. 1073-1080.
6. Makhzani, A. and Frey, B. (2014), "Winner-Take-All Autoencoders", *NIPS'15 Proceedings of the 28th International Conference on Neural Information Processing Systems*, Vol. 2, pp. 2791-2799.
7. Stowell, D. and Turner, R. (2015), Denoising without access to clean data using a partitioned autoencoder. CoRR, abs/1509.05982.

8. Brox T. and Dosovitskiy A. (2016), "Generating Images with Perceptual Similarity Metrics based on Deep Networks", *NIPS'16 Proceedings of the 29th International Conference on Neural Information Processing Systems*, pp. 658-666.
9. Carneiro, G., Garg, R., Kumar, B.G. and Reid, I.D. (2016), "Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue". *European Conference on Computer Vision 2016: Computer Vision – ECCV*, pp 740-756.
10. Bernard, F., Garrido, P., Kim, H., Pérez, P., Tewari, A., Theobalt, C. and Zollhöfer, M. (2017), "MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction", *CoRR*, abs/1703.10580.
11. Dashkevich, A.A. (2016), "Study of multilayer neural networks for automatic feature extraction in solving the problem of pattern recognition. Scientific bulletin of the Tavria agrotechnological state university", *Is. 6, Vol. 2*, pp. 134-139.
12. Dashkevich, A.A. (2017), "Study of convolutional neural networks models for solving the problem of image classification", *Information Processing Systems*, Ivan Kozhedub Kharkiv National Air Force University, *Is. 4 (150)*, pp. 109-112.
13. Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nescic, N., Wang, X., and Westling, P. (2014), "High-resolution stereo datasets with subpixel-accurate ground truth", *GCPR*, Vol. 8753 of *Lecture Notes in Computer Science*, pp 31-42.

Статью представил д-р техн. наук, проф. Національного технічного університету "Харьковский политехнический институт" А.Ю. Ницын

Поступила (received) 24.11.2017

Dashkevich Andrey, Cand. Tech. Sci.
National Technical University "KhPI",
Str. Кургучова, 2, Kharkiv, Ukraine, 61002
Tel.: (095) 388-04-56, e-mail: dashkevich.a@gmail.com
ORCID ID:0000-0002-9963-0998

УДК 004.93

Дослідження моделей згорткових автоенкодерів для виділення ознак в наборах стереозображень / Дашкевич А.О. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2017. – № 50 (1271). – С. 112 – 118.

В роботі проведено процес моделювання навчання без вчителя згорткових автоенкодерів для виділення ознак в наборах стереопар. Досліджено вплив кількості фільтрів, що навчаються і топології автоенкодера на точність відновлення зображень. Визначені конфігурації автоенкодерів з високою якістю відновлення. Ил.: 2. Табл.: 1. Бібліогр.: 13 назв.

Ключові слова: навчання без вчителя; згортковий автоенкодер; виділення ознак; стереопара; відновлення зображень.

УДК 004.93

Исследование моделей сверточных автоэнкодеров для выделения признаков в наборах стереоизображений / Дашкевич А.А. // Вестник НТУ "ХПИ". Серія: Інформатика и моделирование. – Харьков: НТУ "ХПИ". – 2017. – № 50 (1271). – С. 112 – 118.

В работе проведен процесс моделирования обучения без учителя сверточных автоэнкодеров для выделения признаков в наборах стереопар. Исследовано влияние количества обучаемых фильтров и топологии автоэнкодера на точность восстановления изображений. Определены конфигурации автоэнкодеров с высокой точностью восстановления. Ил.: 2. Табл.: 1. Библиогр.: 13 назв.

Ключевые слова: обучение без учителя; сверточный автоэнкодер; выделение признаков; стереопара; восстановление изображений.

UDC 004.93

Study of convolutional autoencoder models for feature extraction from stereo image sets / Dashkevich A.A. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2017. – №. 50 (1271). – P. 112 – 118.

The modeling process of convolutional autoencoder unsupervised learning for feature extraction from stereo images sets is provided. The number of learning filters and topology of autoencoder influence on quality of image reconstruction is researched. Autoencoder configurations of high efficiency reconstruction is defined. Figs.: 2. Tabl.: 1. Refs.: 13 titles.

Keywords: unsupervised learning, convolutional autoencoder; feature extraction; stereo images sets; image reconstruction.