



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 28 за 2021 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

М.Г. Кузьмина

Многослойные
сети-автоэнкодеры в
задачах анализа и
обработки
гиперспектральных
изображений

Рекомендуемая форма библиографической ссылки: Кузьмина М.Г. Многослойные сети-автоэнкодеры в задачах анализа и обработки гиперспектральных изображений // Препринты ИПМ им. М.В.Келдыша. 2021. № 28. 21 с. <https://doi.org/10.20948/prepr-2021-28>
<https://library.keldysh.ru/preprint.asp?id=2021-28>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

М.Г. Кузьмина

**Многослойные сети-автоэнкодеры
в задачах анализа и обработки
гиперспектральных изображений**

Москва — 2021

Кузьмина М.Г. Многослойные сети-автоэнкодеры в задачах анализа и обработки гиперспектральных изображений

Предложена модель пятислойного автоэнкодера, предназначенная для выделения обобщенных «признаков» гиперспектрального изображения и получения сжатого массива данных, определяющих изображение. На первом этапе возможности модели предполагается изучать на основе использования чисто спектральной информации, используя для настройки сети автоэнкодера спектральную функцию ошибок, зависящую от формы спектральных кривых. Планируется испытание автоэнкодера на примерах изображений, содержащихся в трех известных и широко используемых базах гиперспектральных данных (Indian Pines, Pavia University и KSC).

Ключевые слова: Глубокие нейронные сети, многослойные сети-автоэнкодеры, гиперспектральные изображения, обобщенные «признаки» изображения, извлечение «признаков», сжатие исходного массива гиперспектральных данных, определяющих изображение.

Kuzmina M.G. Multilayered autoencoders in problems of hyperspectral image analysis and processing

A model of five-layered autoencoder (stacked autoencoder, SAE) is suggested for deep image features extraction and deriving compressed hyperspectral data set specifying the image. Spectral cost function, dependent on spectral curve forms of hyperspectral image, has been proposed for the autoencoder tuning. At the first step the autoencoder capabilities will be tested based on using pure spectral information contained in image data. The images from well known and widely used hyperspectral databases (Indian Pines, Pavia University и KSC) are planned to be used for the model testing.

Key words: deep neural networks, multilayered autoencoders (stacked autoencoders, SAE), hyperspectral images, image feature extraction, hyperspectral data compression.

1. Введение

1.1. Нейронные сети-автоэнкодеры для получения сокращенной избыточной информации из исходных гиперспектральных данных на основе выделения «признаков» изображения

При восстановлении характеристик земной поверхности по гиперспектральным данным дистанционного зондирования для каждого пикселя требуется произвести анализ соответствующей ему спектральной кривой, определяющей долю отраженного излучения на всем изучаемом интервале спектра. Определение характеристик участка земной поверхности, соответствующего пикселю, основано на сравнении пиксельной спектральной кривой с уже известными спектральными кривыми (из собранных библиотек данных для разных объектов земной поверхности). Трудности анализа гиперспектральных изображений связаны с тремя обстоятельствами: 1) многопиксельными массивами гиперспектральных данных; 2) большим числом спектральных каналов; 3) сложным поведением (с большим числом экстремумов) спектральных кривых. Гиперспектральные изображения обычно содержат десятки тысяч пикселей и сотни спектральных каналов (шириной порядка нескольких нанометров). При этом полные окрестности каждого пикселя, как правило, содержат сильно избыточную информацию. Поэтому при обработке таких изображений возникает необходимость предварительного получения меньших массивов исходных данных с сохранением наиболее существенной информации об изображении.

Таким образом, одним из важных предварительных этапов работы с гиперспектральными изображениями является получение нового, меньшего (сжатого) множества представительных (и избыточных) характеристик изображения. Эта процедура тесно связана с выделением множества обобщенных избыточных характеристик изображения, обычно называемых «признаками» (higher-level features). Получение признаков основано на выявлении и исключении избыточной информации и скрытых корреляций, содержащихся в исходных гиперспектральных данных. При использовании традиционных вычислительных методов обработки многопиксельных изображений одним из проверенных эффективных методов сокращения размерности массива исходных данных является метод главных компонент (Principle Component Analysis, PCA). Другими эффективными и неоднократно опробованными методами снижения размерности массивов больших данных являются нейросетевые подходы, основанные на построении моделей многослойных глубоких нейронных сетей. К такому типу моделей относятся многослойные сети-автоэнкодеры. Однако действие автоэнкодера, которое

иногда можно рассматривать как применение нелинейного обобщения метода PCA, как правило, оказывается более эффективным.

Многослойные сети-автоэнкодеры – глубокие нейронные сети [1] специальной архитектуры. Многослойный автоэнкодер (stacked auto-encoder, SAE) представляет собой специальный вид многослойной сети прямого распространения – многослойный симметричный перцептрон, содержащий несколько внутренних слоев уменьшающегося размера и слой «бутылочная горловина» в середине сети. SAE-сеть производит тождественное преобразование входного слоя на выходной. При этом в результате работы SAE-сети в горловом слое появляется вектор, компонентами которого являются «признаки» – обобщенные характеристики изображения, извлеченные из исходных гиперспектральных данных и содержащие дополнительную существенную и избыточную информацию, определяющую изображение в пространстве меньшей размерности (в так называемом скрытом (latent) пространстве). Получение «признаков» достигается в результате выявления и устранения избыточной информации, а также скрытых взаимосвязей и корреляций в исходных данных. Множество признаков и позволяет затем осуществить операцию сжатия исходных данных и продолжить решение нужной задачи обработки изображения на основе полученных новых данных.

Таким образом, нейросетевые SAE-подходы помимо возможности решения конкретных задач обработки гиперспектральных изображений (задач сегментации и классификации) обладают дополнительными возможностями, к которым можно отнести:

- возможность «извлечения» дополнительной существенной информации – множества глобальных характеристик изображения, в которых устранены избыточная информация и скрытые взаимосвязи и корреляции, содержащиеся в исходных гиперспектральных данных;
- нейросетевой способ сжатия гиперспектральных данных (как правило, без потерь), производимый многослойными сетями-автоэнкодерами, осуществляющими тождественное преобразование входного вектора сети в выходной.

Важным преимуществом автоэнкодеров является также тот факт, что автоэнкодеры являются автоассоциативными сетями. Поэтому их настройка производится посредством самообучения (неконтролируемого обучения, unsupervised learning) и не требует привлечения больших и представительных баз размеченных данных.

Многослойный автоэнкодер впервые был предложен М.Крамером [2] как сеть, предназначенная для сокращения размерности больших массивов данных, представляющих подлежащие обработке изображения. Было показано, что действие автоэнкодера можно интерпретировать как реализацию нелинейной версии классического метода главных компонент (non-linear principle component analysis, NLPCA). При этом, в отличие от PCA, метод NLPCA позволяет

выявить и устранить как линейные, так и нелинейные корреляции в исходных данных.

К настоящему времени разработано несколько типов архитектур глубоких сетей для решения задач анализа и обработки гиперспектральных изображений. К числу сетей часто употребляемых архитектур относятся: обычные многослойные глубокие сверточные сети (DCNN) [1]; сети глубокого доверия (deep belief networks, DBN) [3]; глубокие машины Больцмана (deep Boltzmann machines, DBM) [4]; каскады традиционных сетей-автоэнкодеров (stacked autoencoders, SAE), применяемые для нейросетевого сжатия информации [5, 1]; сети, называемые ограниченными машинами Больцмана (restricted Boltzmann machines, RBM) [6].

1.2. Простые трехслойные автоэнкодеры и метод главных компонент

При использовании традиционных вычислительных методов обработки гиперспектральных изображений для предварительного сжатия размерности массива данных обычно используется классический метод главных компонент (РСА). Как известно, метод РСА состоит в выделении некоррелированных линейных комбинаций в множестве коррелированных исходных гиперспектральных данных, определяющих изображение. Это достигается путем спектрального разложения ковариационной матрицы \hat{C} исходных данных (то есть, представлением \hat{C} в виде суммы ортогональных проекторов на собственные подпространства и в последующем представлении пространства данных в виде суммы взаимно ортогональных собственных подпространств \hat{C}).

Пусть $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathbb{R}^n$ – векторы входных данных, характеризующих анализируемое изображение. Матрицу данных принято записывать в виде

$$\hat{\mathbf{X}} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_m]^T, \quad (1)$$

где каждая строка – вектор обрабатываемых данных, число строк m – количество векторов данных (число пикселей изображения), число столбцов n – размерность пространства данных (число спектральных каналов). Задача определения базиса главных компонент $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ сводится к диагонализации ковариационной матрицы

$$\hat{C} = m^{-1} \sum_{k=1}^m \mathbf{x}_k \mathbf{x}_k^T = \bar{\mathbf{x}} \bar{\mathbf{x}}^T, \quad \bar{\mathbf{x}} = m^{-1} \sum_{k=1}^m \mathbf{x}_k. \quad (2)$$

Ортонормированный базис $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ собственных векторов матрицы \hat{C} , расположенных в порядке убывания собственных значений λ : $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$,

дает решение задачи: коэффициент корреляции между различными координатами равен нулю, и i -я координата в базисе $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ является искомой i -той главной компонентой. Если k – выбранное число главных компонент, которыми желательно ограничиться при анализе изображения ($k < m$), то метод PCA можно интерпретировать как представление матрицы $\hat{\mathbf{X}}$ в виде (см, например, [2]):

$$\hat{\mathbf{X}} = \hat{\mathbf{T}}\hat{\mathbf{P}}^T + \hat{\mathbf{E}}, \quad \hat{\mathbf{T}} = \begin{bmatrix} T_{11} & \dots & T_{1k} \\ \dots & \dots & \dots \\ T_{n1} & \dots & T_{nk} \end{bmatrix}. \quad (3)$$

Матрицу $\hat{\mathbf{P}}$ принято называть матрицей нагрузок (loadings),

$$\hat{\mathbf{P}} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_k] \quad (4)$$

(каждый столбец – вектор главных компонент, k – число столбцов, n – число строк), матрицу $\hat{\mathbf{T}}$ – матрицей отсчетов (scores),

$$\hat{\mathbf{T}} = [T_{ij}], \quad T_{ij} = (\mathbf{x}_i, \mathbf{v}_j) \quad (5)$$

а матрицу $\hat{\mathbf{E}}$ – матрицей ошибок (errors, or residuals): $\hat{\mathbf{E}} = \hat{\mathbf{X}} - \hat{\mathbf{T}}\hat{\mathbf{P}}^T$.

При использовании нейросетевых методов обработки гиперспектральных изображений для предварительного сокращения размерности исходных данных (путем извлечения «признаков») часто используются подходы, основанные на построении многослойных сетей-автоэнкодеров, SAE. Одним из основных достоинств сетей SAE является возможность в задаче извлечения признаков применить методы неконтролируемого глубокого обучения.

Основой для построения всех моделей многослойных автоэнкодеров является модель простого трехслойного автоэнкодера (АЕ). Это сеть прямого распространения с входным и выходным слоями, содержащими одинаковое число нейронов, и единственным внутренним (горловым) слоем, содержащим меньшее число нейронов, чем входной и выходной слои (см. рис. 1).

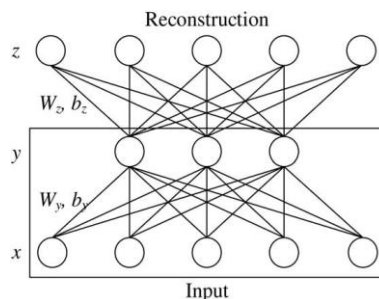


Рис. 1. Схема трехслойного автоэнкодера

Модель простейшего трехслойного АЕ определяется уравнениями (АЕ со «связанными весами»):

$$\begin{aligned} \mathbf{y} &= f(\hat{\mathbf{W}}_y \mathbf{x} + \mathbf{b}_y), \\ \mathbf{z} &= f(\hat{\mathbf{W}}_z \mathbf{y} + \mathbf{b}_z), \quad \hat{\mathbf{W}}_y = \hat{\mathbf{W}}_z^T = \hat{\mathbf{W}}. \end{aligned} \quad (6)$$

Можно считать, что простой автоэнкодер состоит из двух частей: энкодера, осуществляющего сжатие входного вектора (преобразование в «сжатый» вектор меньшей размерности, принадлежащий скрытому пространству «признаков» и представленный функцией кодирования $\mathbf{y} = f(\mathbf{x}; \hat{\mathbf{W}}_y, \mathbf{b}_y)$), и декодера, осуществляющего восстановление входного из «сжатого» и представленного функцией декодирования $\mathbf{z} = f(\mathbf{y}; \hat{\mathbf{W}}_z, \mathbf{b}_z)$.

Настройка АЕ производится путем минимизации функции «ошибки» (cost function) $c(\mathbf{x}, \mathbf{z}) = \text{cost}(\mathbf{x}, \mathbf{z})$, которая может быть определена различными способами. Изменения весов при настройке часто задаются формулами вида

$$\tilde{W} = W - \eta \frac{\delta c(\mathbf{x}, \mathbf{z})}{\delta W}, \quad b_y = b_y - \eta \frac{\delta c(\mathbf{x}, \mathbf{z})}{\delta b_y}, \quad b_z = b_z - \eta \frac{\delta c(\mathbf{x}, \mathbf{z})}{\delta b_z}. \quad (7)$$

Для повышения эффективности работы автоэнкодеров (как однослойных, так и многослойных) часто используется комбинация оправдавших себя «находок»:

1) вместо сигмоидальной функции активации нейронов сети используется функция LReLU (Leaky Rectified Linear Activation)

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ \frac{x}{a} & \text{if } x < 0 \end{cases}$$

где параметр a подбирается для конкретных обрабатываемых данных;

2) используется специальная нормализация исходных входных данных;

3) используются «связанные» веса (в матрицах сетевых связей энкодера $\hat{\mathbf{W}}_y$ и декодера $\hat{\mathbf{W}}_z$);

4) при настройке АЕ используется метод стохастического градиентного спуска (stochastic gradient descent, SGE) с применением методов оптимизации.

После настройки АЕ для последующего решения задач обработки (например, задачи сегментации изображения) слой \mathbf{z} (реконструкция слоя \mathbf{x} , содержащего исходные данные) может быть убран, и дальнейшая обработка осуществляется через скрытый слой \mathbf{y} . Слой \mathbf{y} содержит меньшее множество «извлеченных» в процессе работы автоэнкодера наиболее информативных характеристик изображения. При достаточно точном воспроизведении входных данных (близости \mathbf{z} к \mathbf{x}) слой \mathbf{y} будет содержать существенную сокращенную избыточную информацию, извлеченную из исходных данных \mathbf{x} (из которой устранены корреляции, содержащиеся в данных \mathbf{x}).

Можно отметить соотношение между применением метода PCA и действием однослойного автоэнкодера: в случае использования линейной функции активации $f(x)$ нейронов сети множество компонент вектора y , полученного в горловом слое AE, точно совпадает с числом главных компонент, полученных применением метода PCA к исходным данным x .

Однако в случае нелинейной функции активации имеется преимущество использования автоэнкодеров по сравнению с PCA, связанное с их нелинейностью. Нелинейность позволяет извлекать из исходных данных более существенные обобщенные характеристики изображения, устраняя как линейные, так и нелинейные корреляции в исходных данных (в то время как метод PCA основан на линейных преобразованиях данных и «обрезании» спектра). Кроме того, как оказывается, применение автоэнкодеров позволяет осуществлять обратную реконструкцию исходных данных с меньшей потерей информации.

1.3. Многослойные автоэнкодеры

Многослойные традиционные автоэнкодеры, содержащие несколько внутренних слоев, обладают большими возможностями по сравнению с простыми трехслойными. Пример пятислойного автоэнкодера с тремя внутренними слоями изображен на рис. 2.

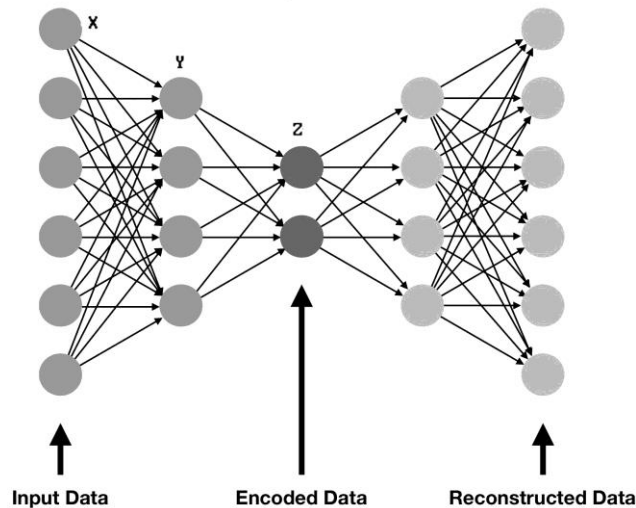


Рис. 2. Пример пятислойного автоэнкодера

Как известно, работу многослойной нейронной сети прямого распространения можно интерпретировать как вычисление определенной композиции многомерных отображений (например, для пятислойного автоэнкодера, изображенного на рис. 2, можно записать $\mathbf{z} = g(f(\mathbf{x}; \theta), \theta = (\{W_{jk}\}, \{g_k\}))$. Путем обучения (настройки) автоэнкодера находятся

значения параметров θ , дающие наилучшую аппроксимацию композиции функций.

Пусть $\hat{\mathbf{X}}$ – $n \times m$ – матрица исходных гиперспектральных данных (1):

$$\hat{\mathbf{X}} = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nm} \end{pmatrix}, \quad \hat{\mathbf{X}} \in \mathbb{R}^{n \times m}, \quad (8)$$

соответствующая изображению (n – число спектральных каналов, m – число пикселей). Используя представление действия метода главных компонент PCA в виде преобразования (3) для матрицы $\hat{\mathbf{X}}$, матрицу $\hat{\mathbf{T}}$, определенную в (5), можно интерпретировать как $(n \times k)$ -матрицу преобразованных данных, $\hat{\mathbf{T}} \in \mathbb{R}^{n \times k}$, а $\hat{\mathbf{P}}$ – как матрицу коэффициентов линейного отображения (проекционного оператора) из \mathbb{R}^n в пространство признаков \mathbb{R}^k ($\hat{\mathbf{P}}\hat{\mathbf{P}}^T = \hat{\mathbf{I}}$). Величина информационных потерь при использовании метода PCA, определяемая матрицей $\hat{\mathbf{E}} = \hat{\mathbf{X}} - \hat{\mathbf{T}}\hat{\mathbf{P}}^T$, тем больше, чем меньше размерность пространства признаков \mathbb{R}^k .

При использовании метода NLPCA преобразование каждого вектора \mathbf{x} , входящего в $\hat{\mathbf{X}}$, из исходного пространства \mathbb{R}^n в пространство признаков \mathbb{R}^k , обобщается (по сравнению с методом PCA) путем введения произвольной нелинейной функции:

$$\mathbf{t} = \mathbf{G}(\mathbf{x}), \quad \text{где} \quad t_j = G_j(\mathbf{x}). \quad (9)$$

При этом в компоненту $G_j(\mathbf{x})$ принято называть j -той нелинейной собственной компонентой вектора \mathbf{x} . Восстановление исходных компонент производится по формулам

$$\tilde{x}_j = F_j(\mathbf{t}), \quad F_j(\mathbf{t}) = G_j^{-1}(\mathbf{t}). \quad (10)$$

Информационные потери при использовании метода NLPCA по-прежнему определяются матрицей $\hat{\mathbf{E}} = \hat{\mathbf{X}} - \hat{\mathbf{T}}\hat{\mathbf{P}}^T$. Выбор функций G и F может определяться, например, из соображений минимизации информационных потерь. Так как метод NLPCA реализуется путем построения модели нейронной сети прямого распространения, функции G и F включают зависимость от функции активации сетевых нейронов.

2. Примеры успешного использования многослойных автоэнкодеров для получения «признаков» изображения и сокращения размерности массива исходных гиперспектральных данных

Отметим некоторые примеры успешного применения SAE-подходов в решении задач анализа и обработки гиперспектральных изображений.

Одним из первых примеров всестороннего изучения возможностей SAE-сетей в задачах анализа гиперспектральных изображений была задача анализа смешанного растительного покрова в двух областях земной поверхности – в районах Kennedy Space Center (KSC), FL, USA и в окрестности города Павия (Pavia, Italy) [7]. Была построена пятислойная модель SAE, состоящая из входного слоя, трех одинаковых слоев меньшего размера и выходного слоя. Выходной слой SAE затем связывается с полносвязным слоевым классификатором, работающим на принципах анализа методом логистической регрессии. В первом случае гиперспектральное изображение содержало 512×614 пикселей с разрешением 18 метров, 176 спектральных интервалов и 13 классов растительности. Во втором случае это было 610×340 - пиксельное изображение, содержащее 115 спектральных интервалов. Для решения задачи классификации гиперспектральных изображений была построена модель SAE-сети, содержащая 4 – 6 скрытых слоев (с одинаковым размером скрытых слоев и числом 20 – 60 нейронов в слое). Таким образом, в целом модель основана на сочетании методов нелинейного обобщения PCA, глубокого обучения и логистической регрессии.

В работе были испытаны возможности SAE-метода при использовании чисто спектральной и объединенной спектрально-пространственной информации, содержащейся в гиперспектральных данных. В последнем случае метод продемонстрировал более высокую точность классификации. Было также показано, что метод, основанный на использовании глубокой SAE-сети, доставляет большую точность классификации по сравнению с такими проверенными методами, как метод PCA и метод опорных векторов (support vector machine, SVM) [8]: точность 98,5% вместо 96,6% для SVM-метода.

Интересные результаты были получены в работах [9, 10], где SAE-автоэнкодер был использован для извлечения обобщенных спектрально-пространственных признаков изображения.

SAE-автоэнкодер, построенный в работе [11], доставляет метод спектрально-пространственной классификации гиперспектральных изображений и позволяет извлекать совокупность спектрально-пространственных признаков изображения. Метод демонстрирует ряд преимуществ по сравнению с другими методами классификации, также основанных на построении моделей глубоких сетей.

Возможности многослойных автоэнкодеров в задаче создания семантически сегментированных карт арктической растительности высокого

разрешения по гиперспектральным спутниковым данным были исследованы в работе [12]. Сети-автоэнкодеры были использованы для неконтролируемого предварительного сжатия исходных гиперспектральных данных. Модель глубокой сверточной сети (DCNN) была далее построена для последующего решения задачи семантической сегментации изображения. При этом был использован кусочный принцип обработки входных изображений (с разделением на фрагменты, содержащие 9, 36 и 144 пикселей). В результате удалось достигнуть точности сегментации 66% - 96%. Было также отмечено, что для увеличения точности семантической сегментации арктического растительного покрова требуются более детальные базы размеченных гиперспектральных данных.

Методы анализа и обработки гиперспектральных изображений, основанные на построении усовершенствованных моделей обычных глубоких сверточных сетей DCNN (не автоэнкодеров) также интенсивно развиваются. При этом используются как подходы, оперирующие с чисто спектральной информацией, содержащейся в гиперспектральных данных [13, 14, 15, 16], так и подходы, позволяющие разными способами включать спектрально-пространственную информацию [9, 10, 17, 18, 19, 20, 21, 22, 23, 24, 25]. Более подробную информацию по методам обработки гиперспектральных изображений, основанным на применении моделей обычных DCNN сетей, можно найти в [26, 27].

Стоит упомянуть интересные примеры моделей пятислойных сетей-автоэнкодеров, которые были построены для решения задачи классификации сортов пшеницы по денситограммам [28, 29]. Денситограмма – одномерное изображение (полоса электрофореграммы после оцифровки), состоящее примерно из 4000 пикселей, позволяющее по спектру определить соответствующий ему белок. При использовании для решения задачи классификации многослойных перцептронов, настраиваемых путем обучения методом ВР (back propagation) для классификации денситограмм, содержащих 50 сортов пшеницы (при размере денситограммы 3225), потребовалась бы настройка более, чем 10^6 параметров сети, что потребовало бы очень длительного обучения сети. Поэтому возникла необходимость в существенном сокращении объема входных данных с сохранением наиболее важной информации. Была построена модель пятислойного автоэнкодера (вида изображенного на рис. 2) с функцией активации LReLU. Оптимальную архитектуру SAE-сети удалось получить применением стохастического градиентного спуска, дополненного методом оптимизации ADAM (Adaptive Moment Estimation) (при подходящей нормировке исходных данных). Оптимальная архитектура имела вид $(L_1, L_2, L_3, L_2, L_1) = (M, 256, 64, 256, M)$, где M – размер денситограммы ($M = 3225$). После сжатия изображения до 64 основных признаков удалось практически точно восстановить его в выходном слое. При обучении SAE (в облачной инфраструктуре с применением карт

GPU) использовался сервис FloydHub (включая TensorFlow и KERAS). В результате общее число настраиваемых параметров автоэнкодера удалось снизить до 315 и достигнуть степени сжатия около 50. После сжатия исходных входных данных была применена глубокая DNN-сеть для классификации.

В работе [16] были построены модели SAE, позволяющие получать множество чисто спектральных обобщенных признаков гиперспектрального изображения и на их основе осуществлять последующее сжатие исходного массива данных. Настройка SAE производилась с применением метода неконтролируемого обучения. Сокращенные данные использовались далее для решения задач кластеризации (сегментации) и классификации гиперспектральных изображений с использованием чисто спектральных характеристик изображений. Особое внимание было уделено выбору функции ошибки (cost, or objective function), минимизируемой при настройке SAE. Функция ошибки строилась зависящей от величины так называемого «спектрального угла» (с добавлением регуляризирующих слагаемых), что позволяло эффективно сравнивать формы двух спектральных кривых. Использование функций ошибки, зависящих от формы спектральных кривых изображения, а также привлечение методов оптимизации при настройке SAE позволило достигнуть высокой точности решения последующей задачи классификации изображений.

Возможности модели тестировались на трех известных базах гиперспектральных изображений – Indian Pines, Pavia University и KSC (Kennedy Space Center) (каждая из баз содержит 10000 изображений). Точность решения задач обработки изображений развитым методом сравнивалась с точностью решения тех же задач такими хорошо известными методами, как PCA, ICA (Independent Component Analysis), метод k – средних, SVM. Как было выяснено, в целом метод, основанный на построении SAE-модели, обеспечивает большую точность решения задач кластеризации и семантической сегментации гиперспектральных изображений, чем перечисленные методы. Основная причина заключается в том, что построенный SAE-автоэнкодер позволяет хорошо изучить пространство «признаков», и поэтому SAE-метод дает возможность четко выделять обобщенные «признаки», отражающие форму спектральных кривых гиперспектральных изображений.

3. Модель пятислойного автоэнкодера для получения сжатого массива гиперспектральных данных на основе использования чисто спектральной информации

Пусть пятислойный автоэнкодер, изображенный на рис. 2, имеет слои с числами нейронов L_j , ($j=1, \dots, 5$), $L_1 = L_5 = 6$, $L_2 = L_4 = 4$, $L_3 = 2$. Его архитектуру удобно характеризовать вектор-строкой

$$(L_1, L_2, L_3, L_4, L_5) = (6, 4, 2, 4, 6).$$

Вектор состояния слоя j , $\mathbf{z}^{(j)} \in \mathbb{R}^{L_j}$, преобразуется в вектор состояния слоя $j+1$, $\mathbf{z}^{(j+1)} \in \mathbb{R}^{L_{j+1}}$ по формуле

$$\mathbf{z}^{(j+1)} = \hat{\mathbf{W}}^{(j)} f(\mathbf{z}^{(j)}) + \mathbf{b}^{(j)},$$

где $\hat{\mathbf{W}}^{(j)}$ – веса $(L_j \times L_{(j+1)})$ –матрицы связей слоев j и $j+1$, f – функция активации нейронов сети.

Принцип работы автоэнкодера — получить в выходном слое сети (в данном случае – в слое L_5) вектор состояния, наиболее близкий к входному. Это достигается самонастройкой автоэнкодера как глубокой сети (с применением неконтролируемого обучения).

При использовании для анализа изображения чисто спектральной информации, содержащейся в гиперспектральных данных (множества спектральных кривых, соответствующих всем пикселям изображения), требуется ввести подходящую характеристику, отражающую различие форм двух спектральных кривых.

Пусть $\mathbf{P} = (P_1, \dots, P_N)$ и $\mathbf{Q} = (Q_1, \dots, Q_N)$ – два спектральных вектора. Следуя [16], используем в качестве (функционального) расстояния между спектральными кривыми величину

$$D_\alpha(\mathbf{P}, \mathbf{Q}) = \sum_n (p_n - q_n)(\log p_n - \log q_n) \equiv \sum_n (p_n \log p_n / q_n + q_n \log q_n / p_n),$$

$$p_n = P_n / \sum_n P_n, \quad q_n = Q_n / \sum_n Q_n \quad (*)$$

Определим автоэнкодер параметрами:

$$(N, M, K, M, N), \text{ т.е. } L_1 = L_5 = N, \quad L_2 = L_4 = M, \quad L_3 = K, \quad (**)$$

где N – число нейронов входного и выходного слоев автоэнкодера (число компонент вектора \mathbf{P} , совпадающее с числом спектральных интервалов), M и K – числа (варьируемые) нейронов во внутренних слоях автоэнкодера ($M > K$). В качестве функции активации нейронов сети можно испытать как сигмоидальную функцию, так и LReLU. Так же как в [16], для настройки параметров автоэнкодера используем функцию оценивания (cost function)

$$E(f(\mathbf{z}^{(L)}), \mathbf{y}) = \sum_n^N [f(\mathbf{z}_n^{(L)}) / \sum_k^N f(\mathbf{z}_k^{(L)}) - \mathbf{y}_n / \sum_k^N \mathbf{y}_k] - \dots - \quad (***)$$

$$- \log f(\mathbf{z}_n^{(L)}) - \log \sum_n^N f(\mathbf{z}_n^{(L)}) - \log \mathbf{y}_n + \log \sum_n^N \mathbf{y}_n$$

Проверенным путем получения оптимальной архитектуры автоэнкодера (см., например, [28, 29]) является применение стохастического градиентного спуска, дополненного методом оптимизации ADAM. Предполагается создание программы, реализующей метод извлечения «признаков» изображения и сжатия гиперспектральных данных, основанный на предложенной версии модели пятислойного автоэнкодера (с использованием открытых библиотечных модулей в Internet и сервиса FloydHub). Проверочные выборки для тестирования метода предполагается брать в трех известных открытых базах гиперспектральных данных - Indian Pines, Pavia University и KSC.

4. Сверточные автоэнкодеры

Сверточные автоэнкодеры (CAE) – многослойные CNN-сети, предназначенные для восстановления на выходе сети 2D изображения, поданного на вход с одновременным извлечением «признаков» изображения. Так же, как в случае традиционных автоэнкодеров архитектура CAE состоит из двух зеркально симметричных частей – кодирующей и декодирующей (восстанавливающей изображение по его сжатой версии).

Кодирующая часть CAE включает сверточные слои и слои пулинга (как правило, используется *max*-pooling). Декодирующая часть содержит слои, осуществляющие операции, обратные свертке и пулингу (см. Рис. 3) и производит реконструкцию исходного изображения по нескольким картам его «признаков».

При работе с CAE-сетями, как с любыми глубокими сетями, используется ряд специальных приемов, ускоряющих процесс длительной настройки глубоких сетей. Сюда относятся: известный метод Dropout (по мере настройки сети часть случайно выбираемых сетевых связей удаляется с некоторой вероятностью), метод добавления шума в веса связей сети (или во входные данные), использование методов частично контролируемого обучения при настройке сети, использование контролируемого прерывания процесса преднастройки. Используются также методы оптимизации при применении метода стохастического градиентного спуска (SGD), в том числе, популярный метод ADAM.

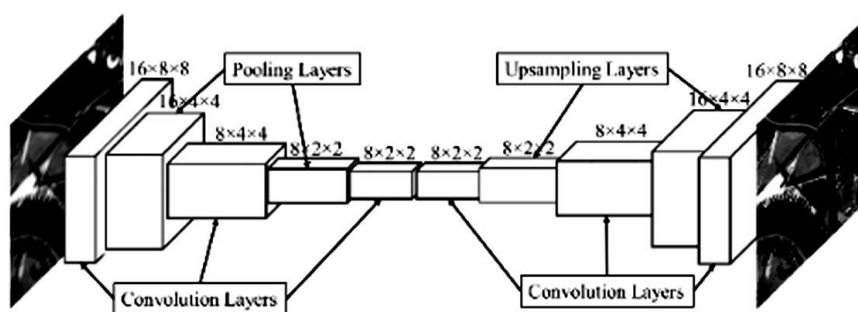


Рис. 3. Архитектура сверточного автоэнкодера

Большое число параметров САЕ-сети требует представительного множества обучающих данных. Поэтому обычно встает вопрос о пополнении баз данных, требующихся для настройки САЕ. Кроме того, как правило, приходится прибегать к предварительной и последующей (более точной) настройке САЕ-сети. (Предварительная настройка нужна для формирования бассейнов притяжения динамической системы, соответствующей САЕ-сети, и играет роль регуляризации).

Основное достоинство САЕ-сетей в том, что они сохраняют пространственную информацию, содержащуюся в гиперспектральных данных. Кроме того, САЕ-сети оказываются хорошо приспособленными для иерархического извлечения обобщенных глубоких «признаков» гиперспектральных изображений. При получении «признаков» слои *max-pooling* выполняют функции «регуляризаторов».

Целый ряд сложных задач обработки гиперспектральных изображений удалось решить методами, основанными на построении моделей сверточных автоэнкодеров с обеспечением качественных результатов. Можно, например, отметить успешное решение задачи сегментации изображений на основе построения моделей сверточных автоэнкодеров, настраиваемых методами неконтролируемого обучения [30, 31, 32, 33, 34, 35]. При этом методы, основанные на моделях САЕ, позволяют предварительно извлекать большее число (спектрально-пространственных) «признаков» изображения по сравнению с методами, основанными на моделях традиционных автоэнкодеров (SAE) [36].

В последнее время активно развиваются модели так называемых трехмерных сверточных нейронных сетей (3D CNN), в том числе, и модели 3D сверточных автоэнкодеров [32]. Область применения – разнообразные задачи анализа и обработки больших данных. Помимо задач анализа гиперспектральных изображений сюда относятся задачи анализа медицинских снимков, связанные, например, с распознаванием опухолей.

Так, для классификации гиперспектральных изображений очень высокого разрешения (VHR HIS) весьма эффективным оказался метод, основанный на модели сети архитектуры энкодер-декодер, предложенный в работе [37]. Это

трехмерная сверточная сеть, позволяющая одновременно учесть спектрально-пространственную и контекстуальную информацию, заключенную в данных, определяющих изображение. Показано, что точность классификации значительно возрастает по сравнению с точностью метода классификации, основанного на модели обычной глубокой сверточной сети (на 7.42 %).

В рамках подхода 3D CNN в работе [38] развит метод выделения спектрально-пространственных «признаков» гиперспектральных изображений с последующей классификацией изображений при использовании неконтролируемого обучения. Для достижения более высокой точности классификации использована фильтрация Габора в пространственной области. Показано, что метод спектрально-пространственной классификации, основанный на построении моделей 3D CNN, превосходит по точности классификации аналогичный метод, основанный на построении моделей обычных глубоких сверточных сетей (2D CNN).

5. Заключительные замечания

Предложена модель пятислойного автоэнкодера, содержащего три внутренних слоя, допускающего оптимизацию архитектуры (вариацию числа нейронов во внутренних слоях и функции активации нейронов сети). Настройка сети автоэнкодера будет производиться с использованием спектральной функции ошибок, зависящей от формы спектральных кривых гиперспектрального изображения.

Предполагается создание программы (с привлечением открытых библиотечных модулей в Internet и сервиса FloydHub), реализующей метод извлечения обобщенных признаков анализируемого изображения, сжатия исходных гиперспектральных данных и последующего решения задач сегментации и классификации изображения на основе полученного «сжатого» множества данных, определяющего изображение.

Проверочные выборки для тестирования метода предполагается брать в трех известных открытых базах гиперспектральных данных – Indian Pines, Pavia University и KSC.

Список литературы

1. I. Goodfellow, Y. Bengio, A. Courville, Deep Learning. The MIT Press, 2016. ISBN: 0262035618.
2. M.A Kramer. Nonlinear principal component analysis using autoassociative neural networks. AICHE Journal, vol. 37, no. 2, pp. 233- 243, 1991.
3. G. E. Hinton, S. Osindero, and Y. Teh, A fast learning algorithm for deep belief nets, Neural Computation, vol. 18, no. 7, pp. 1527–1554, 2006.
4. R. Salakhutdinov, and G.E. Hinton. Deep Boltzmann machines, in Proc. Int. Conf. Artif. Intell. Statist., Clearwater Beach, FL, USA, pp. 448–455, 2009.
5. Y. Bengio., P. Lamblin, D. Popovici, and H. Larochelle. Greedy layer-wise training of deep networks, in Proc. Neural Inf. Process. Syst., Cambridge, MA, USA, pp. 153–160, 2007.
6. G. Hinton et al., A practical guide to training restricted Boltzmann machines, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. UTM TR2010-003, 2010.
7. Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu. Deep learning-based classification of hyperspectral data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, v. 7, no. 6., pp. 2094–2107, 2014.
8. N. Cristianini, N. Shawe-Taylor. An Introduction to Support Vector Machines and other kernel-based learning methods / Cambridge University Press, 2020. [SVM].
9. Y. Chen, X. Zhao, and X. Jia, Spectra-spatial classification of hyperspectral data based on deep belief network, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 6, pp. 2381–2392, 2015.
10. Y. Chen, Z. Lin, X. Zhao, X. Jia. Spectral–spatial classification of hyperspectral data based on deep belief network, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1939-1404, 2017. [DOI: [10.1117/1.JRS.11.042604](https://doi.org/10.1117/1.JRS.11.042604)]
11. G. Abdi, F. Samadzadegan, P. Reinartz, Spectral–spatial feature learning for hyperspectral imagery classification using deep stacked sparse autoencoder, J. Appl. Remote Sens. 11(4), 042604 (2017), doi: 10.1117/1.JRS.11.042604.

12. Z. L. Langford, J. Kumar, and F.M. Hoffman. Convolutional neural network approach for mapping Arctic vegetation using multi-sensor remote sensing fusion. IEEE International Conference on Data Mining Workshops, 2017. DOI [10.1109/ICDMW.2017.48](https://doi.org/10.1109/ICDMW.2017.48)
13. A.Krizhevsky, I.Sutskever, and G.Hinton. Imagenet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105
14. F. Hu, G.-S. Xia, J. Hu, and L. Zhang. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens 7, 14680-14707, 2015; doi:10.3390/rs71114680
15. K.Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, arXiv: 1409.1556v6 [cs.CV] 10 Apr 2015.
16. L.Windrim, .Ramakrishnan, A.Melkumyan, R.J.Murphy, A.Chlingaryan. Unsupervised feature-learning for hyperspectral data with autoencoders. Remote Sens. 2019, 11, 864; <https://doi.org/10.3390/rs11070864>
17. Z. Lin, Y. Chen, X. Zhao, Spectral-spatial classification of hyperspectral image using autoencoders, Conference Paper December 2015. <https://www.researchgate.net/publication/271492851>
18. Jun Yue, Wenzhi Zhao, Shanjun Mao & Hui Liu, Spectral–spatial classification of hyperspectral images using deep convolutional neural networks// Remote Sensing Letters, v. 6, pp. 468-477, 2015.
19. A. Santara, K. Mani, P. Hatwar, A. Singh, A. Garg, K. Padia, and P. Mitra, Bass net: Band-Adaptive Spectral-Spatial feature learning neural network for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing, v. 55, n. 9, pp. 5293–5301, 2017.
20. N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, Deep learning classification of land cover and crop types using remote sensing data // IEEE Geoscience and Remote Sensing Letters, DOI:10.1109/LGRS. 2017. 2681128.
21. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. Manzagol, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, Journal of Machine Learning Research, vol. 11, pp. 3371–3408, 2010.

22. W. Zhao and S. Du, Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach, *IEEE Transactions on Geoscience and Remote Sensing*, v. 54, n. 8, pp. 4544–4554, 2016.
23. A. Romero, C. Gatta, and G. Camps-Valls, Unsupervised deep feature extraction for remote sensing image classification // *IEEE Transactions on Geoscience and Remote Sensing*, v. 54, n. 3, pp. 1349–1362, 2016.
24. P. Zhong, Z. Gong. A hybrid DBN and CRF model for spectral-spatial classification of hyperspectral images // *Stat., Optim. Inf. Comput.*, v. 5, pp 75–98. 2017.
25. L. Mou, P. Ghamisi, and X. X. Zhu, Unsupervised spectral-spatial feature learning via deep residual conv-deconv network for hyperspectral image classification, *IEEE TGRS*, vol. 56, no. 1, pp. 391–406, Jan 2018.
26. Л.П. Басс, М.Г.Кузьмина, О.В.Николаева. Сверточные нейронные сети с глубоким обучением в задачах обработки гиперспектральных спутниковых данных // *Препринты ИПМ им.М.В.Келдыша*, № 282, 2018.
27. Л.П. Басс, М.Г. Кузьмина, О.В. Николаева, Ю.А. Пластинин, И.Ю. Скрыбышева. Применение глубоких сверточных сетей в задачах обработки гиперспектральных спутниковых данных, Моногр. «Искусственный интеллект в космической технике: состояние, перспективы применения». Под науч. ред. докт. техн. наук А.Н. Балухто, 2021.
28. Г.А.Ососков, П.В.Гончаров, А.В. Цитринов. Применение глубоких нейронных сетей для классификации изображений. 19 Международная научно-техническая конференция «Нейроинформатика-2017», 2-6 окт. 2017, НИЯУ МИФИ, Москва.
29. G. Ososkov, P. Goncharov. Shallow and deep learning for image classification. *Opt. Mem. Neural Netw.* 26(4), 221–248, 2017.
30. J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction, *ICANN 2011, Part I, LNCS 6791*, pp. 52–59, 2011. (Springer-Verlag, 2011)
31. V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for Image Segmentation, arXiv: 1511. 00561v3 [cs.CV] 10 Oct 2016.

32. J. Penttilä. A method for anomaly detection in hyperspectral images, using deep convolutional autoencoders // Master's Thesis in Information Technology, 2017, University of Jyväskylä.
33. X. Guo, X. Liu, E. Zhu, and J. Yin, Deep clustering with convolutional autoencoders, in Proc. ICONIP . Springer, 2017, pp. 373–382.
34. J. Nalepa, M. Myller, Y. Imai, Ken-ichi Honda, T. Takeda, and M. Antoniak, Unsupervised segmentation of hyperspectral images using 3D convolutional autoencoders, arXiv: 1907.08870v1 [cs.CV] 20 Jul 2019.
35. A. Azarang, H.E. Manoochehri, and N. Kentarnavaz, Convolutional autoencoder-based multispectral image fusion, IEEE Access, v. 7, pp. 35673-35683, 2019.
36. B. Ghazanfari, F. Afghah, Multi-level feature learning on embedding layer of convolutional autoencoders and deep inverse feature learning for image clustering, arXiv: 2010.02343v1 iv: 5 Oct 2020.
37. Z. Zhang, T. Jiang, C. Liu, and L. Zhang, An effective classification method for hyperspectral image with very high resolution based on encoder-decoder architecture, Selected Topics in Applied Earth Observations and Remote Sensing, pp (99): 1-1, 2020. DOI: [10.1109/JSTARS.2020.3046245](https://doi.org/10.1109/JSTARS.2020.3046245)
38. B. Praveen, and V. Menon, Study of Spatial–Spectral Feature Extraction Frameworks With 3-D Convolutional Neural Network for Robust Hyperspectral Imagery Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, 2021.

Благодарности:

Автор благодарен А. Минкину за ссылки на работы по SAE 2019, а также за ссылки на сайты в Internet, содержащие готовые версии программ, где использованы SAE-методы.

Простые версии SAE можно найти на сайте:

<https://neurohive.io/ru/osnovy-data-science/avtojenkoder-tipy-arhitektur-i-primenenie/>

Некоторые готовые реализации автоэнкодеров:

1. Простой автоэнкодер, реализованный на Pytorch:

https://github.com/avijit9/Contractive_Autoencoder_in_Pytorch

2. deephyp: deep learning for hyperspectral imagery (на Tensorflow):

https://github.com/avijit9/Contractive_Autoencoder_in_Pytorch

Примеры гиперспектральных изображений (снимки Sentinel2, включающие 12 спектральных каналов) содержатся на сайте:

<https://scihub.copernicus.eu/dhus/#/home>

Оглавление

1. Введение	3
1.1. Нейронные сети-автоэнкодеры для получения сокращенной неизбыточной информации из исходных гиперспектральных данных на основе выделения «признаков» изображения	3
1.2. Простые трехслойные автоэнкодеры и метод главных компонент	5
1.3. Многослойные автоэнкодеры	8
2. Примеры успешного использования многослойных автоэнкодеров для получения «признаков» изображения и сокращения размерности массива исходных гиперспектральных данных.....	10
3. Модель пятислойного автоэнкодера для получения сжатого массива гиперспектральных данных на основе использования чисто спектральной информации	12
4. Сверточные автоэнкодеры.....	14
5. Заключительные замечания.....	16
Список литературы.....	17