

Рис. 1 – Общая структура модели

Такая модель успешно используется в задачах машинного перевода и подписи изображений. Для обучения такой модели необходим набор данных, содержащий пары звук-текст [1]. Обучение модели разделено на следующие шаги:

1. обучение системы распознавания речи (кодировщика), используя звук-текст обучающие примеры;
2. удаление Connection temporal classification и Softmax слоев нейронной сети;
3. блокировка всех слоев кодирующей нейронной сети чтобы избежать их переобучения;
4. подключение к модели синтеза голоса (декодера);
5. обучение конечной модели с использованием только звука в качестве входных и выходных данных.

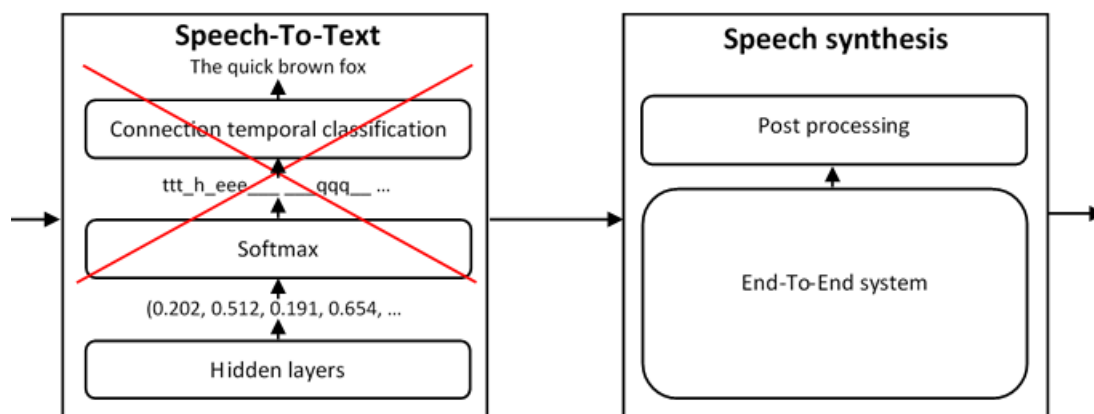


Рис. 2 – Подробная структура модели

В качестве кодировщика может быть использована предобученная нейронная сеть, такая как Speech-to-Text-WaveNet, DeepSpeech. Нейронная сеть для синтеза голоса должна быть обучена с нуля из-за нестандартного представления входных данных. Такое представление данных может содержать больше информации об интонации. Кодировочная нейронная сеть может быть использована для любых примеров, однако декодирующая должна быть дообучена в отдельности для синтеза конкретного голоса [2].

Список использованных источников:

1. AaronvandenOord, SanderDieleman, HeigaZen, KarenSimonyan, OriolVinyals, AlexGraves, NalKalchbrenner, AndrewSenior, KorayKavukcuoglu, WaveNet: AGenerativeModelforRawAudio [Электронный ресурс]. – 2016. – Режим доступа: <https://arxiv.org/abs/1609.03499>.
2. Wei Ping, Kainan Peng, Andrew Gibiansky, Sercan O. Arık, Ajay Kannan, Sharan Narang, Deep Voice 3: Scaling Text-to-Speech with Convolutional Sequence Learning [Электронныйресурс]. – 2018. – Режимдоступа: <https://arxiv.org/abs/1710.07654>.

ПОВЫШЕНИЕ КАЧЕСТВА ИЗОБРАЖЕНИЯ НА БАЗЕ АЛГОРИТМОВ НЕЙРОННЫХ СЕТЕЙ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Никитин Г.Ю.

Петровский А.А. – д.т.н., профессор

В настоящее время остро стоит проблема наличия различных шумов и артефактов на изображениях, что приводит к сильному снижению их качества. На практике наиболее распространённые шумы колеблются от аддитивного шума до мультипликативного шума. Такая деградация может оказать значительное влияние на точность методов компьютерной обработки в медицине, сделать анализ и распознавание изображений трудными и ненадёжными.

Существует большое множество алгоритмов, способных убрать с изображения шумы. В последнее время было проведено много исследований алгоритмов шумоподавления в вейвлет области, например, был предложен алгоритм на основе вейвлет преобразования с последующей отсечкой шумовых коэффициентов [1], или, к примеру, алгоритм оценки шумовых вейвлет коэффициентов, основывающийся на моделях

распределения Гаусса и Лапласа [2]. Также одним из алгоритмов, позволяющих убрать с изображения различные шумы, является алгоритм обработки изображения нейронной сетью в вейвлет области [3, 4].

Вейвлет преобразование представляет собой перевод изображения из временного представления в пространственно-временное представление. В данном случае рассматривается дискретное вейвлет преобразование (discretewavelettransform, DWT). Каждая часть такого вейвлет преобразования подаётся на свою нейронную сеть для обработки и сборки в обесшумленное изображение при помощи обратного вейвлет преобразования (inversediscretewavelettransform, IDWT). Структура алгоритма представлена на рисунке 1:

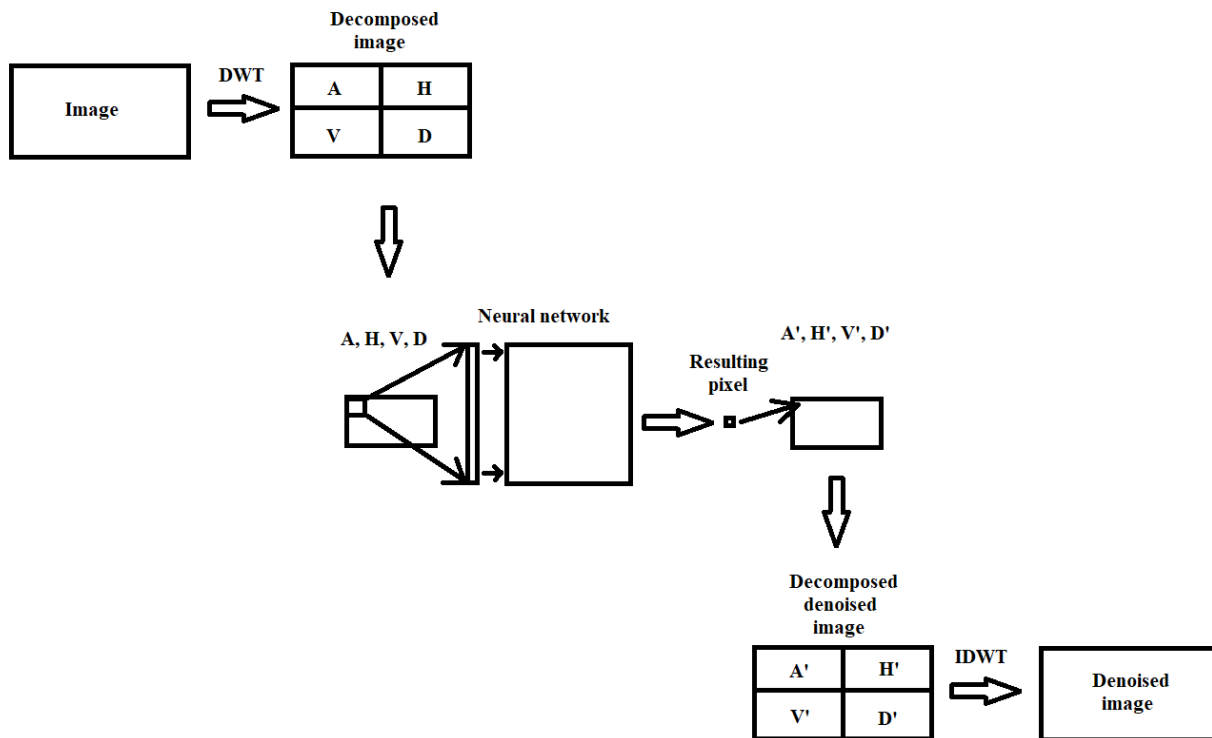


Рис. 1 – Структура алгоритма.

Как уже было сказано, в начале работы изображение проходит вейвлет преобразование (отмечено DWT). Затем каждая часть преобразованного изображения (отмечено A, H, V, D) подаётся на соответствующую нейронную сеть. При этом на обработку нейронным сетям (Neuralnetwork) подаются области некоторого размера (в данной работе был использован размер области 5x5 пикселей) с шагом в 1 пиксель. В результате обработки каждой области нейронная сеть возвращает уже обесшумленный пиксель (Resultingpixel). После завершения обработки всех областей полученные пиксели собираются вместе и выполняется обратное вейвлет преобразование (IDWT) и получение исходного изображения.

Нейронные сети представляют из себя двухслойные линейный нейронные сети, с 10 нейронами на 1 слое и 1 нейроном на 2 (выходном) слое. Обучение проводилось методом Левенберга-Марквардта.

Алгоритм был протестирован на нескольких тестовых изображениях с шумами (изображения не участвовали в выборке для обучения нейронных сетей). Результаты работы алгоритма представлены на рисунках 2а, 2б и 2в. Значения PSNRпредставлены в таблице 1.



Рис. 2а – Исходное изображение без шума.

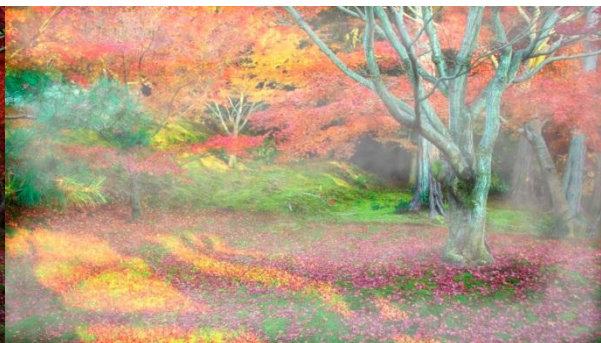


Рис. 2б – Исходное изображение с шумом.



Рис. 2в – Изображение с шумом после обработки алгоритмом.

Табл. 1 – Значения PSNR для обрабатываемых изображений.

Изображение	PSNR (до обработки)	PSNR (после обработки)
Для обучения нейронных сетей	8.676	21.031
Тестовое	9.853	20.017
Тестовое	10.415	20.835

Таким образом была представлена система шумоподавления, использующая вейвлет преобразование и линейные нейронные сети. Нейронные сети ищут зависимость между исходными и зашумленными коэффициентами вейвлет преобразования, что позволяет эффективно подавлять различные шумы на изображениях.

Основные преимущества такого подхода:

- Обработка изображения в вейвлет области занимает значительно меньше времени, чем при обработке исходного изображения из-за возможности параллельной обработки нескольких областей вейвлет преобразования одновременно;
- Вышеописанное свойство также работает и для обучения нейронных сетей, что позволяет значительно сократить время работы.

Основной недостаток – данный алгоритм (как и многие, если не все) не даёт 100% шумоподавления и восстановления исходного изображения. Это объясняется тем, что часть информации теряется необратимо и лишь часть можно восстановить при использовании алгоритмов восстановления и шумоподавления.

Список использованных источников:

1. Donoho D., Adapting to unknown smoothness via wavelet shrinkage / D. Donoho, I. Johnstone // *Biometrika* – 1994.
2. Yu H., Image denoising using trivariate shrinkage filter in the wavelet domain and joint bilateral filter in the spatial domain / H. Yu, L. Zhao, H. Wang // *IEEE Trans. Image Proc.* – 2009.
3. Zhang S., Image denoising using a neural network based non-linear filter in wavelet domain / S. Zhang, E. Salari
4. Bhutata G., Image enhancement by wavelet-based thresholding neural network with adaptive learning rate / Bhutata G., Anand R., Saxena S. // *IET Image processing* – 2010.

СИСТЕМА МОДЕЛИРОВАНИЯ ЭФФЕКТА ПОТЕРИ СЛУХА ПО АУДИОГРАММЕ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Порхун М.И.

Вашкевич М. И. – кандидат технических наук, доцент

Предлагается система моделирования эффекта потери слуха по аудиограмме. Задача моделирования потери слуха сводится к обработке сигнала таким образом, чтобы человек с нормальным слухом услышал звук так, как его воспринимает тугоухий человек. Метод заключается в субполосной обработке сигнала с применением компрессоров динамического диапазона (КДД). Настройка компрессоров выполняется исходя из аудиограммы тугоухого человека.

Введение. Несмотря на то, что за последние годы были разработаны различные системы моделирования потери слуха, ни одна из них не получила широкого распространения. Методы моделирования потери слуха многими специалистами не воспринимается всерьёз, хотя их применение позволяет обеспечить реалистичную демонстрацию как коммуникативных, так и психосоциальных эффектов нарушения слуха. Основная цель моделирования потери слуха заключается в том, чтобы показать людям с