РЕЦЕНЗИЯ
на выпускную квалификационную работу

Петряшина Ильи Евгеньевича
на тему:
«Разработка нейросетевого метода распознавания речи в условиях сильной зашумленности»

Распознавание речи в условиях сильной зашумленности является одной из ключевых задач и имеет широкое применение во множестве областей. Учитывая, что существующие модели часто показывают низкое качество в условиях шума, особенно в реальном времени, данное исследование крайне актуально. В работе данная задача разбивается на 2 подзадачи, которые рассматриваются отдельно – задачу детекции голосовой активности в реальном времени и задачу распознавания речи в условиях сильной зашумленности.

Для детекции голосовой активности в реальном времени в работе предложен алгоритм RealVADR, использующий существующие нейросетевые модели детекции. Рассмотрено, как параметры данного алгоритма влияют на качество распознавания речи. Для распознавания речи в условиях сильной зашумленности в работе предложен оригинальный подход DASR, суть которого заключается в совместном использовании моделей расшумления речи и моделей распознавания. Подход позволяет дообучать модели расшумления, пропуская градиент ошибки непосредственно с выхода моделей распознавания речи. Таким образом модели расшумления оказываются подстроены именно под модель распознавания.

Сильной стороной работы является обширное тестирование предложенных методов. Про оба алгоритма показано, что их применение показывает сопоставимое с существующими техниками качество решения непосредственной задачи, что подтверждает: внедрённые для улучшения применимости метода доработки не навредили его исходной цели. Метод DASR оказался даже лучше существующих техник с точки зрения кросс-доменной генерализации.

По-личному интересной находкой Ильи Евгеньевича для меня является обнаружение пользы и даже длины интервалов, которые вместе с детектированным участком речи следует включать в распознавание. Действительно, предоставив распознавателю речи информацию о том, как среда звучала до начала речи, помогает улучшить качество распознавания речи.

Актуальность исследования не вызывает сомнений, ведь наличие шума сильно затрудняет распознавание речи, и предложенные методы направлены на решение этой до сих пор остро стоящей проблемы. Гибкость и обобщаемость предложенного подхода, а также его применимость к различным речевым доменам, свидетельствуют о высокой прикладной пользе предложенного метода.

К недостаткам работы можно отнести не слишком большую репрезентативность экспериментов в разделе 5.2.7 и 5.2.10, что может ограничивать общую применимость выводов в этих частях. Изображения подписаны сжато, что принуждает читателя хорошо задуматься об их посыле. Так, неясно, что кодируется красным, а что зелёным цветом на Рисунке 1, и что же отражают пунктирные линии на Рисунке 8. Эти недостатки не несут критического значения, и не искажают безусловно положительного впечатления от работы.

Мне было бы интересно узнать, какие именно характеристики отличают "по-новому" расшумленные записи от таких, которые получаются существующими

методами. Я бы рекомендовал провести отдельный пласт экспериментов, посвящённый именно такому анализу. Устранилась ли проблема с удалением частотных полос, описанная на странице 28 (раздел 5.2.3)? Какие части слов или, может, фонемы стали распознаваться по-другому?

Я считаю, что работа полностью соответствует уровню ВКР магистра и заслуживает оценки «отлично».

Рекомендую присвоить выпускнику квалификацию магистра

Рецензент:
Сериков Олег Алексеевич,
программист-разработчик
лаборатории нейронных систем
и глубокого обучения МФТИ

(подпись рецензента, заверенная по месту работы)

20 июня 2024

Подпись руки
заверяю:
зав канцелярией
административного отдела
М.А. Гусева

КАНЦЕЛЯРИЯ