

Система виділення людського голосу з зашумленого аудіозапису з використанням глибокого навчання

Виконав: Харченко Дмитро Олександрович

Науковий керівник:

к.т.н. доц. Дідковська Марина Віталіївна



Актуальність роботи

- Очищення шуму в онлайн розмові
- Створення контенту без професійного обладнання
- Системи голосової аутентифікації

- Об'єкт дослідження – виділення людського голосу.
- Предмет дослідження – глибоке навчання для обробки шуму в аудіо.
- Мета – дослідження можливості використання глибокого навчання для задач фільтрації шуму на звукозаписі.

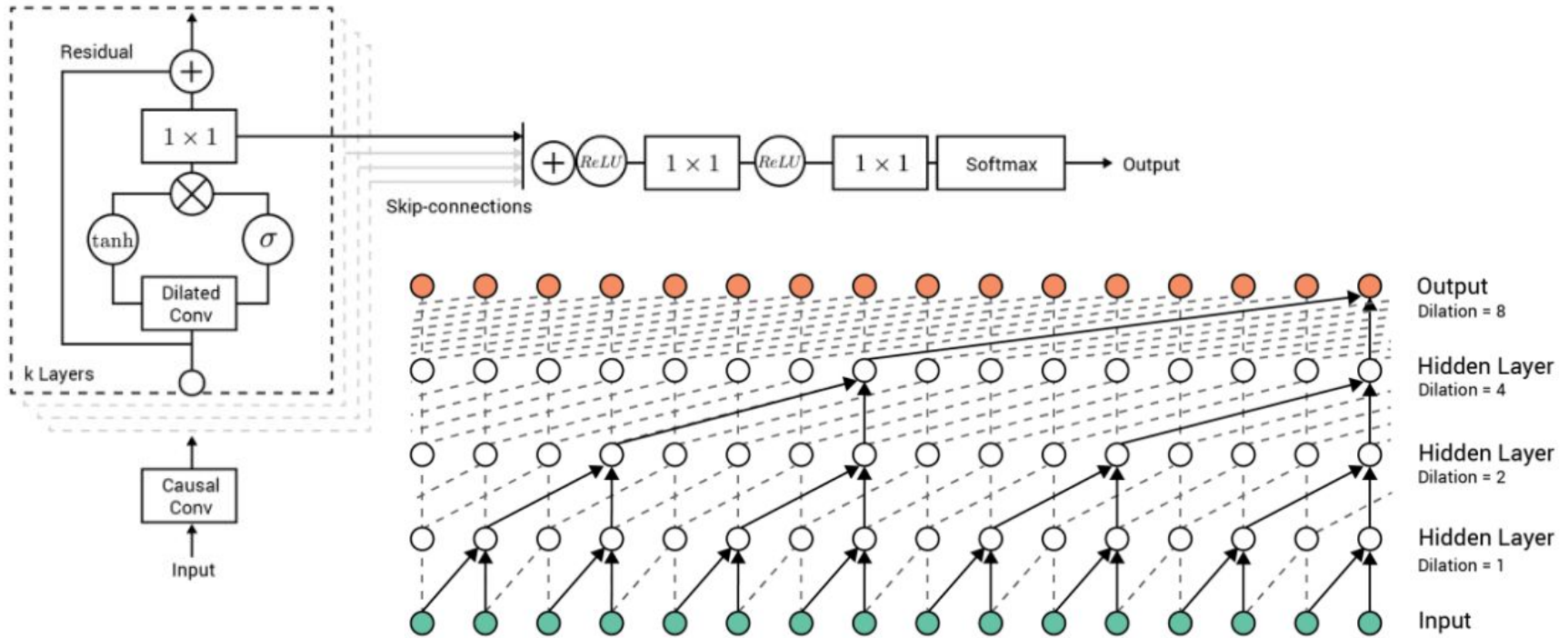
Постановка задачі

- Проаналізувати існуючі підходи та методи фільтрації шуму на звукозаписі.
- Розробити архітектуру фільтрування звуку в реальному часі.
- Реалізувати модуль розробленої системи.
- Сформулювати висновки щодо роботи реалізованого модуля.

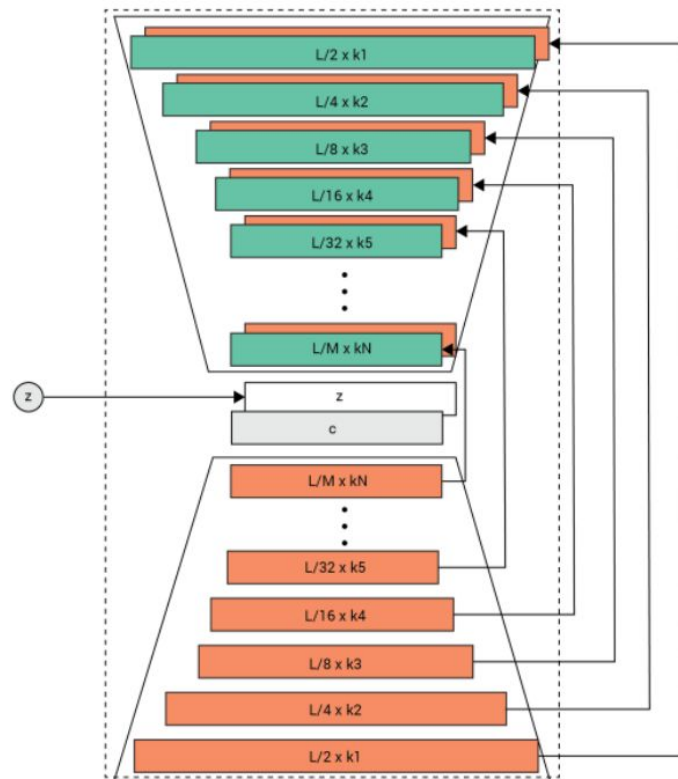
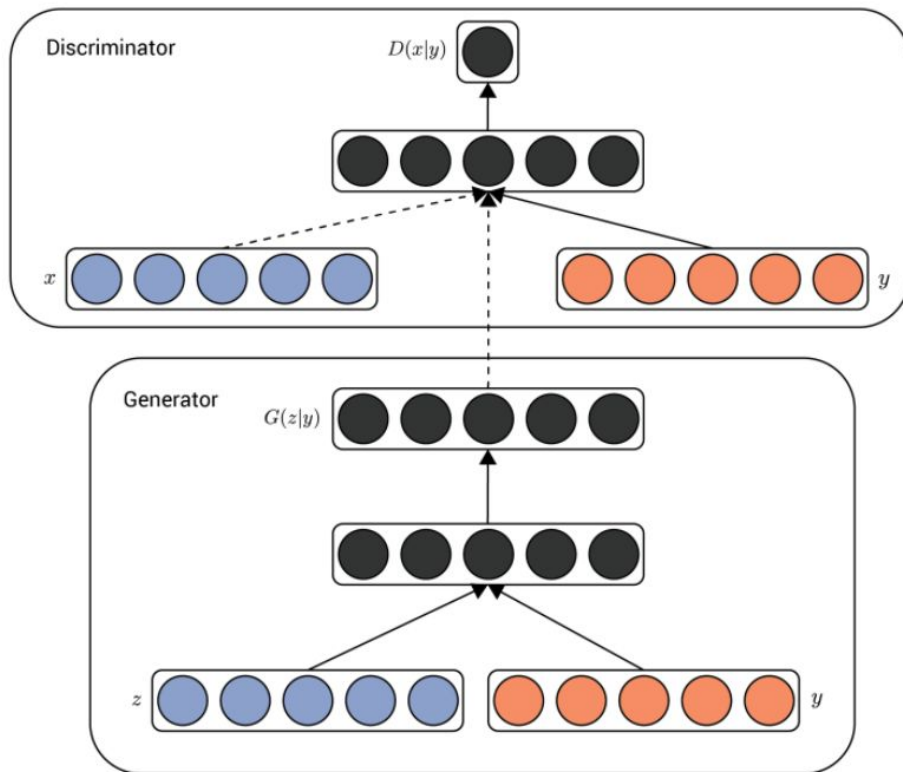
Існуючі методи

- Класичні:
 - Фільтр Вінера
 - Перетворення Фур'є
- Глибоке навчання:
 - Генеративно змагальні мережі - SEGAN
 - Згорткові авторегресори - WaveNet denoising
 - Згортково рекурентна мережа - EHNet

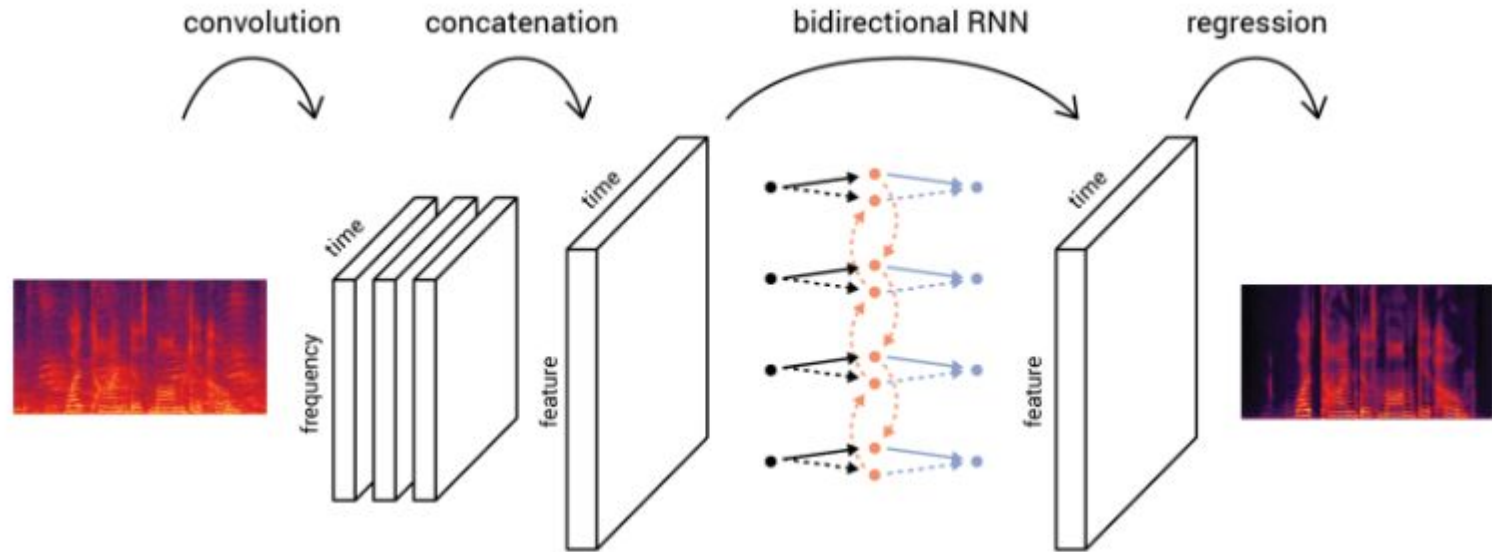
Архітектура WaveNet




Архітектура SEGAN



Архітектура EHNet

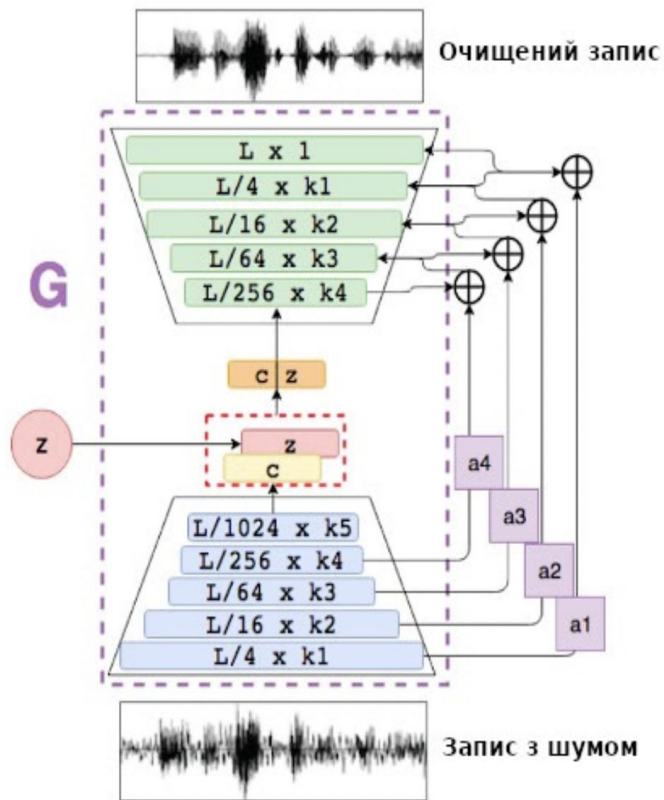




Критерії порівняння

- Якості:
 - PESQ
 - STOI
 - Людська оцінка
- Швидкості:
 - Час обробки звукозапису

Запропонована архітектура

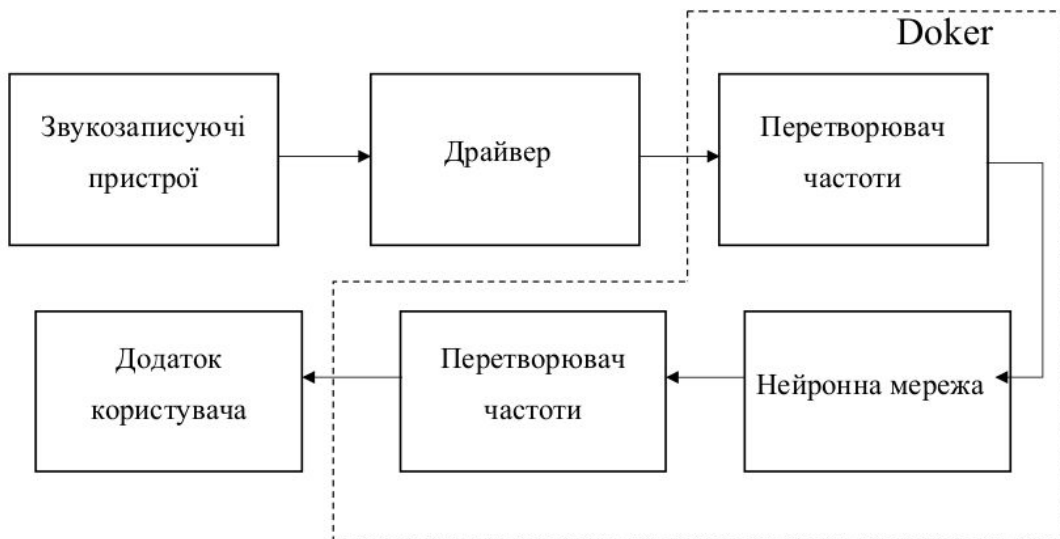


Зміни:

- менша кількість шарів
- ваги у обхідних з'єднаннях

Очікувані переваги:

- швидкість роботи
- прийнятна якість

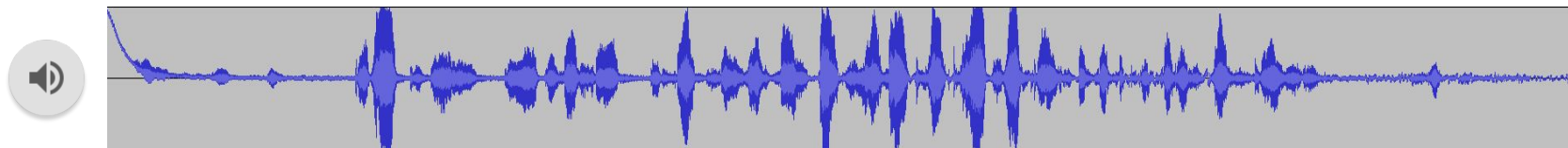


Аналіз результатів

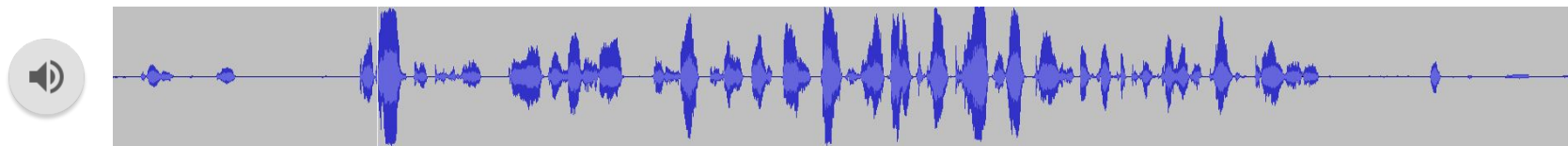
Модель	Час роботи	PESQ	STOI	Людська оцінка
WaveNet denoising	1.7с	2.91	-	3.01
SEGAN	2.9с	3.22	0.932	3.47
Модифікований SEGAN	1.8с	3.10	0.906	3.30
EHNet	5.1с	2.71	0.886	2.87
Wiener	-	3.02	0.921	3.00
Noisy	-	3.01	0.921	3.07

Результати роботи

Запис з шумом



Запис після обробки



Висновки

- Запропонована модифікація архітектури мережі для прискорення при збереженні якості та реалізована у якості модуля системи виділення голосу.
- Проведено порівняння з існуючими методами, який показав що запропонована архітектура має кращу якість за WaveNet при аналогічній швидкості роботи.
- Опубліковано статтю, що зображає запропоновану модифікацію.

Рекомендації для подальших досліджень

- Дослідження роботи системи при використанні української мови
- Покращення якості оцінки результатів
- Покращення якості, в умовах існування RTX Voice

Публікації

Харченко Д. Пришвидшена генеративно змагальна мережа для відділення голосу від шуму на звукозаписі. *Наука онлайн: Міжнародний електронний науковий журнал*. 2020. №5. URL: <https://nauka-online.com/ua/publications/tehnicheskie-nauki/2020/5/prishvidshena-generativno-zmagalna-merezha-dlya-viddilennya-golosu-vid-shumu-na-zvukozapisi/>

Дякую за увагу!

