Технічні науки

УДК 004.934.5

Харченко Дмитро Олександрович

студент

Навчально-наукового комплексу

«Інститут прикладного системного аналізу»

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

ПРИШВИДШЕНА ГЕНЕРАТИВНО ЗМАГАЛЬНА МЕРЕЖА ДЛЯ ВІДДІЛЕННЯ ГОЛОСУ ВІД ШУМУ НА ЗВУКОЗАПИСІ

Анотація. *У статті розглянута модифікація генеративно змагальної мережі для покращення звукозапису, зроблені висновки про якість роботи мережі та проведено порівняння з оригінальною мережею та іншими методами вирішення задачі очищення звукозапису від шуму.*

Ключові слова: *генеративно змагальні нейронні мережі, згорткові нейронні мережі, глибоке навчання, очищення звукозапису, фільтрація шуму*

Для виділення голосу на звукозаписі традиційно використовували методи з галузі обробки сигналів та методи які застосовують теорію підпросторів. Однак, в останні роки широкого розповсюдження здобули методи машинного навчання, зокрема глибоке навчання та нейронні мережі. Вони розв'язують ті задачі, де традиційні математичні рішення вимагають довготривалої розробки або взагалі неможливі. Такі задачі зазвичай визначаються як пошук закономірностей, який саме й виконується при машинному навчанні. Задача виділення голосу з запису може бути сформульована як пошук закономірностей у звуковій хвилі голосу та її відмінності від хвилі шуму. Таким чином вирішення цієї задачі за допомогою машинного навчання є цілком природнім.

Остання розробка RTX Voice від Nvidia показала, що існує значний попит на технологію фільтрації шуму в реальному часі. Проте згадана розробка значно обмежена цільовими користувачами, оскільки вона працює лише на операційній системі Windows та графічному прискорювачі від Nvidia. Як наслідок, проблема фільтрації шуму у реальному часі є актуальною, має як попит так і доказ можливості його вирішення.

Серед сучасних розробок у галузі глибокого навчання, для задачі фільтрації звукозапису, виділяються наступні 3 архітектури:

- SEGAN [1]
- WaveNet denoising [2]
- EHNet [3]

SEGAN це генеративно змагальна мережа пристосована до генерації очищеного запису. WaveNet представляє собою мережу що поєднує ідеї авторегресії та згорткової мережі і вирішує задачі генерування звукозапису з мовою. WaveNet denoising є модифікацією оригінальної мережі, що налаштована саме для очищення запису від шумів. EHNet в свою чергу використовує згрорткові та рекурентні мережі для обробки звуку.

Серед згаданих архітектур, найкраще за якістю очищення працює SEGAN, проте вона значно поступається WaveNet denoising у швидкості роботи.

Таким чином в даній роботі розглядається модифікація для прискорення мережі з найкращою якістю.

Архітектура генеративно змагальної мережі поділяється на генератор та оцінювач. Генератор створює новий звукозапис використовуючи оригінал, в той же час коли оцінювач намагається визначити вірогідність того що зразок звукозапису є оригіналом, а не згенерованим. Архітектура генератора нагадую автоенкодер, а саме використовує згортки та обернені згортки для виділення ознак запису та реконструкції нового запису за цими

ознаками. Архітектура оцінювача повторює згорткову частину архітектури генератора, що завершується класифікатором за виділеними ознаками.

Після навчання, для вирішення задачі використовується якраз генератор. Тому саме архітектура генератора нас цікавить найбільше. Генератор оригінального SEGAN зображено на рисунку 1. В оригінальній роботі архітектура має 11 шарів згортки та 11 шарів оберненої згортки.

Паскаль та інші у своїй роботі [4] використовують модифікований SEGAN у задачі генерування звичайної мови з шепоту. Їх модифікація полягає у зменшенні кількості шарів за допомогою збільшення кроку згортки та введенні вагів у обхідні зв'язки між шарами згортки та оберненої згортки. Іншою модифікацією була зміна норми регуляції, щоб прибрати обмеження на амплітуду вихідного сигналу. Остання модифікація нам не цікава оскільки в їх задачі вихідна амплітуда має бути більшою за вхідну, що не потрібно в задачі фільтрування. Проте перші дві зміни нас зацікавили, оскільки, зменшення кількості шарів значно пришвидшує мережу, в той час коли додаткові ваги обхідних з'єднань дозволяють зменшити втрати якості пов'язані зі зменшенням кількості шарів. Запропонована архітектура генератора зображена на рисунку 2.

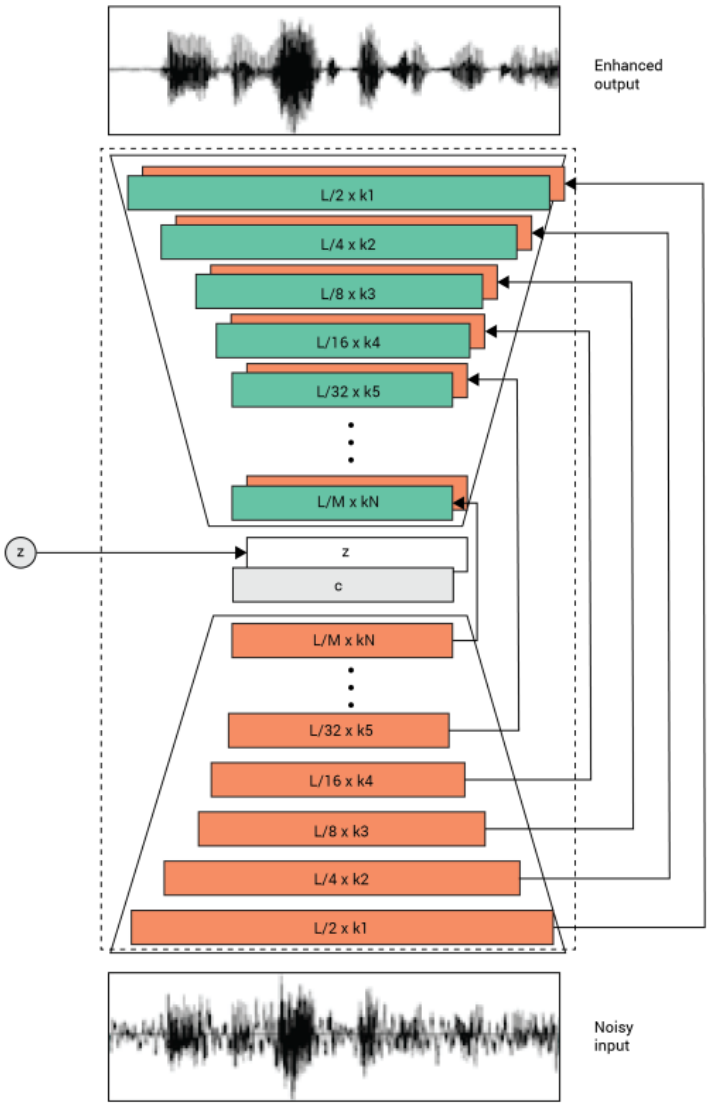

Рис. 1. Архітектура генератора SEGAN [1]

Для порівняння роботи запропонованої модифікації, вона тренувалась разом з мережами запропонованими в роботах [1-3] на одному наборі даних.

Аналогічно до цих робіт в якості оцінки роботи мережі використовується оцінка PESQ, STOI та людська оцінка за шкалою від 1 до 5. Результати оцінювання якості наведено в таблиці 1.

В якості порівняння з традиційними методами використаємо фільтр Віньера [5] як відправну точку. Результати оцінки зображено в таблиці 1. Для оцінки швидкості береться час очищення запису довжиною 2с.

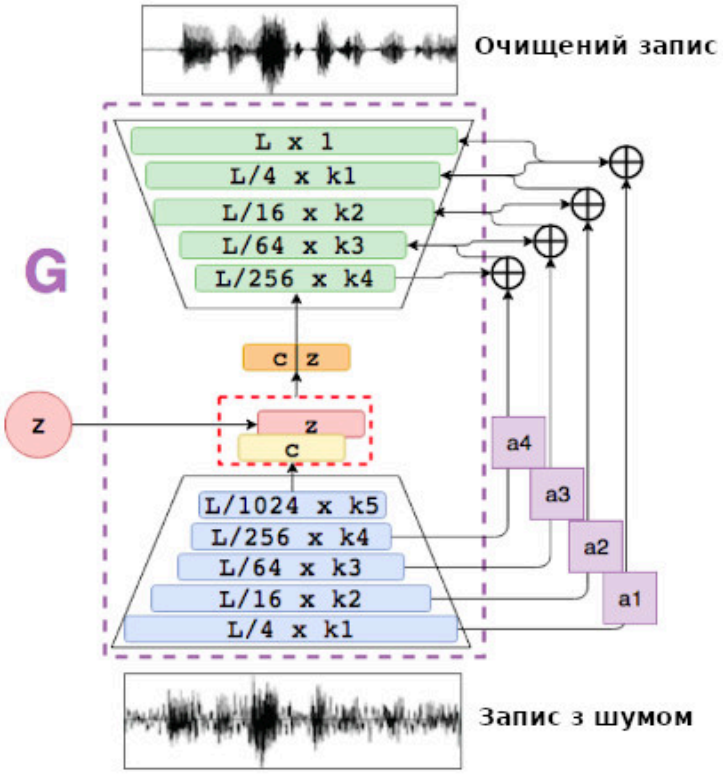

Рис. 2. Швидка модифікація SEGAN

Таблиця 1

Оцінка очищення запису різними мережами

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

Оцінка STOI відсутня для WaveNet через особливості цієї архітектури, а саме те, що вона використовує майбутні дані, що призводить до незначного зменшення вихідного запису.

Як показали результати, запропонована модифікація все ж програє оригінальній SEGAN за якістю, проте випереджає WaveNet. Водночас час

роботи запропонованої мережі майже вдвічі менший за оригінальну та порівняний з часом роботи WaveNet. Таким чином запропонована мережа виконує поставлене завдання, а саме значно прискорює оригінал при незначних втратах якості.

Параметри системи на якій проводилось тестування:

- Графічний прискорювач Nvidia GeForce 1050Ti
- Центральний процесор – Intel Core-i5 2500k
- Операційна система – Ubuntu 18.04 LTS
- Python 3.7
- TensorFlow 1.15
- CUDA 10.1

Література

1. Pascual S., Bonafonte A., and SerraÃÄ J. SEGAN: Speech Enhancement Generative Adversarial Network. 2017.
2. Rethage D., Pons J., and Serra X.. A Wavenet for Speech Denoising. 2018.
3. Zhao H. et al. Convolutional-Recurrent Neural Networks for Speech Enhancement. 2018.
4. Pascual S. et al. Whispered-to-voiced Alaryngeal Speech Conversion with Generative Adversarial Networks. 2018.
5. Esfandiar Zavarehei. Wiener Filter. 2019. URL:https://mathworks.com/matlabcentral/fileexchange/7673-wiener-filter.