

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»
Кафедра інформатики факультету інформатики

**АЛЬТЕРНАТИВНИЙ МЕТОД КОДУВАННЯ ТА ДЕКОДУВАННЯ МОВИ
ЗГОРТКОВОЮ НЕЙРОННОЮ МЕРЕЖЕЮ**

Текстова частина до курсової роботи
за спеціальністю „Інженерія програмного забезпечення” 121

Керівник курсової роботи
д.т.н., доц. Глибовець А.М.

(підпис)
“ ” 2020 р.

Виконав студент
Редчиць Є.В.

“ ” 2020 р.

Київ 2020

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»

Кафедра інформатики факультету інформатики

ЗАТВЕРДЖУЮ

Зав. кафедри інформатики

к.ф.-м.н., доц.

_____ Гороховський С.С.

(підпис)

“ ____ ” _____ 2020 р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ
на курсову роботу

студенту Редчицю Є.В. факультету інформатики 1 курсу МП

ТЕМА Альтернативний метод кодування та декодування мови згортковою нейронною мережею

Вихідні дані:

Зміст ТЧ до курсової роботи:

Індивідуальне завдання

Вступ

1 Огляд типової згорткової нейронної мережі

2 Побудова альтернативного методу кодування

3 Реалізація альтернативного методу кодування з модифікацією

Висновки

Список літератури

Дата видачі “ ____ ” _____ 2020 р. Керівник _____

(підпис)

Завдання отримав _____

(підпис)

Тема: Альтернативний метод кодування та декодування мови згортковою нейронною мережею

Календарний план виконання роботи:

№ п/п	Назва етапу дипломного проекту (роботи)	Термін виконання етапу	Примітка
1.	Отримання завдання на курсову роботу.	02.11.2019	
2.	Огляд технічної літератури за темою роботи.	15.11.2019	
3.	Виконати аналіз сучасних методів кодування мовного сигналу	25.11.2019	
3.	Аналіз алгоритмів розпізнавання зображень звичайною та згортковою нейронними мережами	30.12.2020	
6.	Виконання порівняльного аналізу результатів кодування типових методів кодування мовного сигналу з альтернативним методом та аналіз запропонованої модифікації	30.03.2020	
7.	Написання пояснювальної роботи.	20.04.2020	
8.	Створення слайдів для доповіді та написання доповіді.	22.04.2020	
8.	Аналіз отриманих результатів з керівником, написання доповіді	25.04.2020	
10.	Коригування роботи за результатами попереднього огляду роботи.	01.05.2020	
11.	Остаточне оформлення пояснювальної роботи та слайдів.	03.05.2020	
12.	Захист курсової роботи	15.05.2020	

Студент Редчиць Є.В.

Керівник Глибовець А.М.

“ ____ ” _____ 2020 р

Зміст

Анотація	5
Вступ	6
1 ОГЛЯД ТИПОВОЇ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ	8
1.1 Огляд архітектури звичайної нейронної мережі	8
1.2 Приклад побудови ЗНМ	10
1.3 Аналіз шарів ЗНМ	12
2 ПОБУДОВА АЛЬТЕРНАТИВНОГО МЕТОДУ КОДУВАННЯ	17
2.1 Загальна архітектура альтернативної ЗНМ	17
2.2 Зворотне поширення помилки	17
2.3 Автокодер	18
3 РЕАЛІЗАЦІЯ АЛЬТЕРНАТИВНОГО МЕТОДУ КОДУВАННЯ З МОДИФІКАЦІЄЮ	20
3.1 Частотно-часова згортка	20
3.2 Структура альтернативної ЗНМ	23
3.3 Результати дослідження альтернативного методу кодування	24
3.4 Модифікація альтернативного методу кодування мови	26
Висновки	30
Список літератури	31

Анотація

У даній курсовій роботі розглянуто загальні відомості про застосування згоркової нейронної мережі. Також проаналізовано альтернативний метод кодування та декодування мови з використанням спектрограми аудіосигналу та запропоновано модифікацію до альтернативного методу кодування та декодування мовного сигналу з використанням бітового потоку згортковою нейронною мережею.

Ключові слова: згорткова нейронна мережа, спектрограма, аудіосигнал, бітовий потік, кодування, декодування.

Вступ

Актуальність теми. Процес кодування та декодування мови є невід’ємною частиною процесу комунікації між людьми у всіх сферах її діяльності. За умов всесвітньої пандемії коронавірусу (COVID-19) питання спілкування людей на відстані є запорукою їх здоров’я та життя. На сьогоднішній день розроблена велика кількість алгоритмів кодування, які є основою для більшості стандартів, таких як LPC-10, MELP, CELP, G.732 тощо. В залежності від сфери використання кожен з цих стандартів може бути адаптований до відповідних умов та задовольняти потреби користувачів. Але процес вдосконалення алгоритмів не припиняється й досі, так як в різних каналах зв’язку можуть утворюватися низка чинників, які впливають на якість мовного сигналу. З моменту використання нейронні мережі для промислових потреб, з’явилися нові можливості вдосконалення обробки мови. Це дозволяє позбутися проблем, які притаманні для класичних систем кодування, таких як випадання слів під час розмови, сторонні звуки, шум, нерозбірливість мови тощо.

Структура зорової кори у мозку ссавців була моделлю для створення ЗНМ [1]. Локальне розташування пікселів визначає розпізнавання форми об’єкта, тому ЗНМ починає аналізувати зображення з меншими локальними візерунками, а потім поєднує їх у більш складні форми. Ефективність ЗНМ в розпізнаванні об’єктів на зображеннях може бути основою для вирішення проблеми розпізнавання мовлення. Таким чином, запропонований метод використовує кодування звуку за допомогою зображень [2]

Мета дослідження. Проаналізувати альтернативний метод кодування та декодування мови використовуючи зображення бітового потоку мовного сигналу згортковою нейронною мережею (ЗНМ).

Об’єкт дослідження. Кодування та декодування мови ЗНМ.

Предмет дослідження. Основні алгоритми даних методів, структури вхідних даних, основні переваги і способи використання.

Джерела дослідження. Електронні ресурси.

1 ОГЛЯД ТИПОВОЇ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ

ЗНМ дуже схожі на звичайні нейронні мережі: вони складаються з нейронів, які характеризуються ваговими значеннями та вектором зсуву. Кожен нейрон приймає значення, отримує скалярний добуток та за необхідності визначається нелінійною системою. Вся мережа виражає єдину диференційовану функцію оцінок: від необроблених пікселів зображення на одному кінці до оцінок класу на іншому [3].

Архітектура ЗНМ чітко визначає входи як зображення, що дозволяє кодувати певні властивості в архітектурі та робить функцію передачі більш ефективною для реалізації та значно скорочує кількість параметрів у мережі [3].

1.1 Огляд архітектури звичайної нейронної мережі

Нейронні мережі приймають значення вектору і перетворюють його через ряд прихованих шарів. Кожен прихований шар складається з набору нейронів де кожен нейрон повністю з'єднаний з усіма нейронами попереднього шару і де нейрони в одному шарі функціонують повністю незалежно і не поділяють жодних зв'язків. Останній повноз'єднаний шар (ПШ) називається вихідним шаром (ВШ) і в налаштуваннях класифікації він представляє собою оцінку класів [3].

Звичайні нейронні мережі не масштабуються до повних зображень. У CIFAR-10 (Canadian Institute for Advanced Research) зображення мають розмір 32 на 32 на 3 (32 ширина, 32 висота та 3 кольорових канали), тому один повністю пов'язаний нейрон у першому прихованому шарі звичайної нейронної мережі матиме 32 на 32 на 3 дорівнює 3072 вагового значення. Візуальний приклад звичайної нейронної мережі можна побачити на рисунку 1.1 [3].

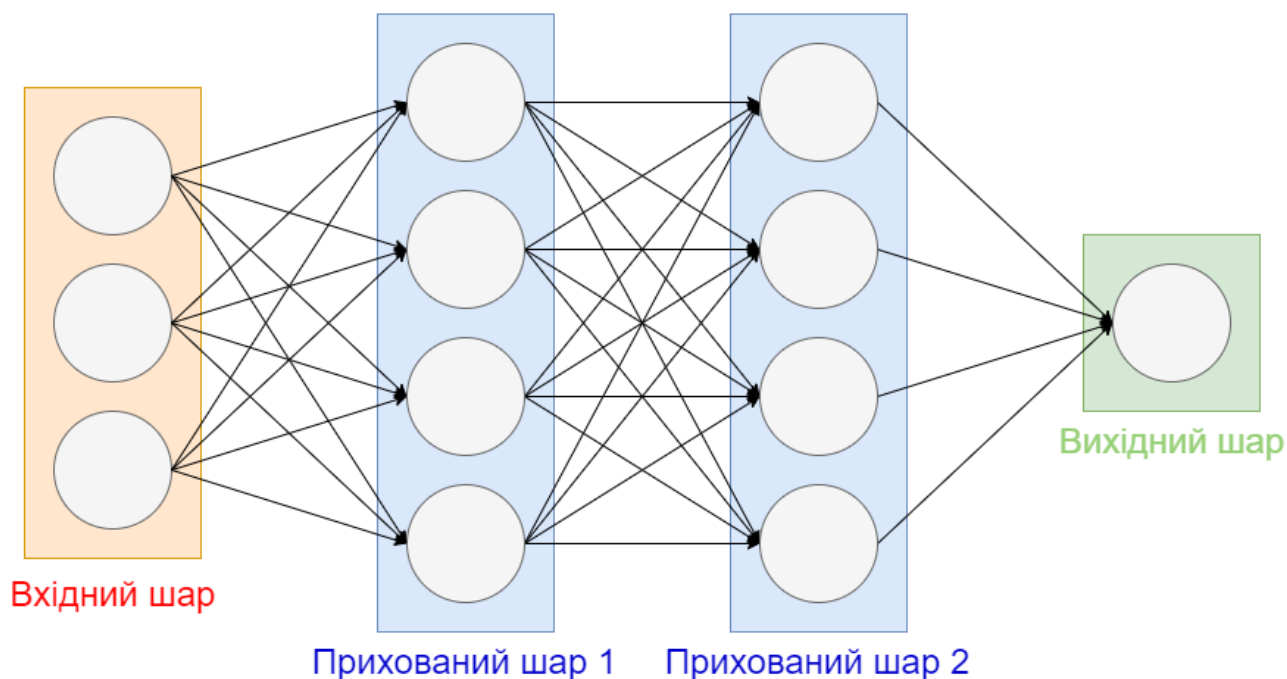


Рисунок 1.1 – Звичайна нейронна мережа

ЗНМ мають перевагу в тому, що вхід складається з зображень і вони описують архітектуру більш раціонально. Зокрема, на відміну від звичайної нейронної мережі, шари ЗНМ мають нейрони, розташовані в 3 вимірах: ширина, висота, глибина. Наприклад, вхідні зображення в CIFAR-10 є вхідними значеннями активації і дані значення мають розміри 32 на 32 на 3 (ширина, висота, глибина відповідно). Нейрони в шарі пов'язані лише з невеликою областю шару перед ним, замість усіх нейронів повністю пов'язаним чином. Кінцевий вихідний шар для CIFAR-10 мав би розміри 1 на 1 на 10, в результаті чого повне зображення буде зменшено в один векторний показник класу, розміщений вздовж глибинного виміру. Приклад ЗНМ зображено на рисунку 1.2 [3].

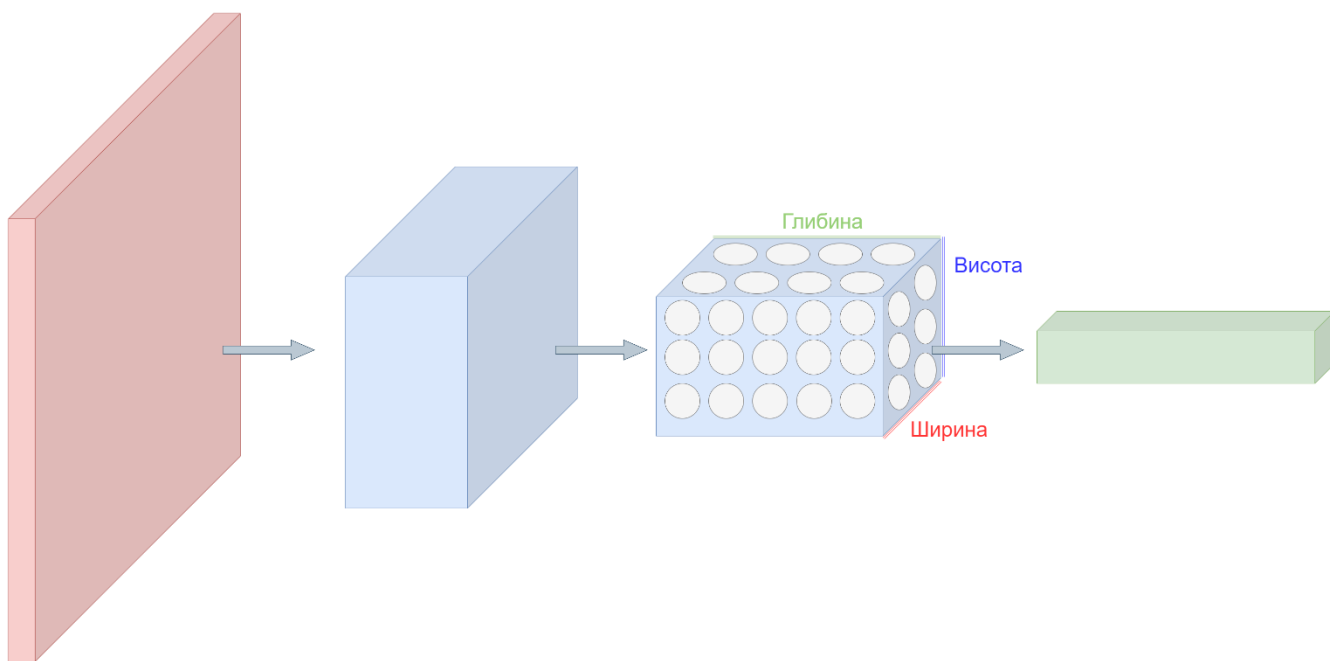


Рисунок 1.2 – Звичайна ЗНМ

1.2 Приклад побудови ЗНМ

ЗНМ – це послідовність шарів де кожен шар перетворює одне значення активацій в інший за допомогою диференційованої функції. Для цього використовуються три основні типи шарів: згортковий шар (ЗШ), агрегувальний шар (АШ) та повноз'єднаний шар (ПШ). Дані шари складаються для того щоб сформувати повну архітектуру ЗНМ [3].

Проста класифікація ЗНМ для CIFAR-10 може мати архітектуру зображену на рисунку 1.3.

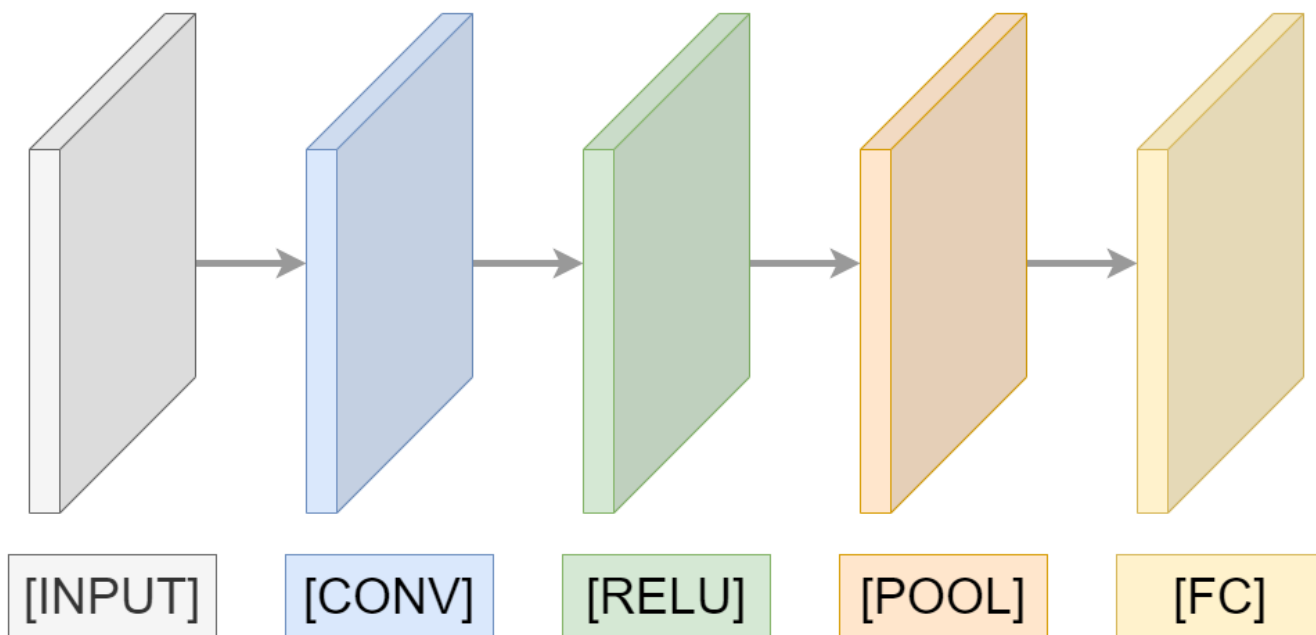


Рисунок 1.3 – Проста класифікація ЗНМ для CIFAR-10

Шар INPUT (32 на 32 на 3) буде містити значення необроблених пікселів зображення, у цьому випадку зображення має характеристики з шириною 32, висотою 32 та трьома кольоровими каналами R, G, B [3].

Шар CONV обчислює вихід нейронів, які з'єднані з локальними регіонами на вході, кожен з яких обчислює скалярний добуток між їх ваговими значеннями та невеликою областю, до якої вони підключені у вхідному значенні. якщо буде прийнято рішення використовувати 12 фільтрів результатом цього буде значення 32 на 32 на 12 [3].

Шар RELU застосує функцію активації елементів, наприклад операція максимізації ($MAX(0, x)$) з нульовим пороговим значенням. Таким чином, значення розміру залишається незмінним, тобто 32 на 32 на 12 [3].

Шар POOL буде виконувати операцію децимації вздовж просторових розмірів (ширина, висота), в результаті чого буде досягнуто значення 16 на 16 на 12 [3].

Шар FC обчислює оцінку класів, в результаті отримуючи значення розміру 1 на 1 на 10, де кожне з 10 чисел відповідає оцінці класу. Як і у звичайних нейронних мережах кожен нейрон у цьому шарі буде з'єднаний з усіма числами попереднього значення [3].

ЗНМ перетворює вихідний шар за шаром від початкових значень пікселів до кінцевих показників класу. Деякі шари містять параметри, а інші не містять. Зокрема, шари CONV/FC виконують перетворення, які є функцією не тільки активації вхідного значення, але і параметрів (вагове значення та вектор зсуву). З іншого боку, шари RELU/POOL будуть реалізовувати фіксовану функцію. Параметри шарів CONV/FC будуть навчатися з градієнтним спуском так, щоб оцінка класу, що обчислює ЗНМ, відповідали міткам у навчальному наборі для кожного зображення [3].

1.3 Аналіз шарів ЗНМ

Згортковий шар – це основний будівельний блок ЗНМ, який виконує більшість складних обчислювальних обрахунків [2, 3].

Параметри ЗШ складаються з набору визначених фільтрів. Кожен фільтр просторово невеликий (по ширині та висоті), але поширюється на повну глибину вхідного значення. Наприклад, типовий фільтр на першому шарі ЗНМ може мати розмір 5 на 5 на 3. Під час проходження вперед кожен фільтр просувається по ширині та висоті вхідного значення та обчислюється скалярний добуток між записами фільтра та входом у будь-якому положенні. Коли фільтр зсувається по ширині та висоті вхідного значення, створюється двовимірний карта активації, яка дає результати цього фільтра у кожному просторовому положенні. Далі мережа вивчить фільтри, які активуються, коли вони бачать певний тип візуальної функції, наприклад, край певної орієнтації чи пляма якогось кольору на першому шарі або цілі візерунки на

більш високих шарах мережі. Кожен фільтр з набору фільтрів у кожному шарі ЗНМ створить окрему двовимірну карту активації. Дані карти активації розташовуються по глибинному виміру і виробляють вихідне значення шару [2, 3].

Глибина ЗШ залежить від кількості різних функцій, використаних у зображенні, отриманих за кількістю фільтрів. Кожен нейрон в ЗШ використовує вагове значення рівно одного фільтра ЗШ. Багато нейронів використовують аналогічні вагові значення. Нейрони в ЗШ сегментуються за допомогою фільтра, який вони використовують. Рецептивне поле нейрона в ЗШ вказує локальну область зв'язку нейрона на попередній шар. Всі нейрони в межах одного ЗШ мають однаковий розмір своїх сприйнятливих полів. Всі зв'язки даного нейрона утворюють поле прийому цього нейрона. Значення сприйнятливого поля завжди дорівнює розміру фільтра певного ЗШ помноженому на глибину попереднього шару [2, 3].

Активация обчислюється шляхом застосування функції активації над потенціалом. В якості функції активації в більшості випадків використовується рампова функція, яка також називається блоком RELU. Якщо конкретна особливість має значення в одній частині зображення, зазвичай ця особливість важлива і в іншій частині зображення. Один з найважливіших гіпер параметрів поруч із кількістю фільтрів – це крок. Визначаючи крок, враховується розмір сприйнятливого поля нейрона та розмір зображення. У разі неправильного вибору слід змінити крок або застосувати нульове накладення, щоб зберегти певний розмір введення або нормалізувати зображення різної форми [2, 3].

Загальноприйнятою практикою є використання **агрегувального шару** (АШ) між послідовними ЗШ в ЗНМ. Його функція полягає в поступовому зменшенні просторових розмірів представлення для зменшення кількості параметрів і обчислень в мережі та для контролю над переповненням. АШ працює незалежно на кожному фрагменті глибини входу і змінює його просторово, використовуючи операцію максимізації. Найпоширенішою формою є шар об'єднання з фільтрами

розміром 2 на 2, нанесеним з кроком по 2 відліки. Кожна операція максимізації у цьому випадку повинна мати максимум 4 числа. Розмір глибини залишається незмінним [2, 3].

АШ приймає значення розміром визначеним в формулі 1.1.

$$W_1 \times H_1 \times D_1 \quad (1.1)$$

де W_1 – ширина;

H_1 – висота;

D_1 – глибина.

АШ потребує два гіпер параметри, а саме просторове значення F , та значення кроку S . АШ генерує значення розміру зазначене в формулі 1.2 [2, 3].

$$W_2 \times H_2 \times D_2 \quad (1.2)$$

де $W_2 = (W_1 - F)/S + 1$;

$H_2 = (H_1 - F)/S + 1$;

$D_2 = D_1$.

На практиці найчастіше зустрічається два варіанти АШ, це АШ з F , що дорівнює 3 з S , що дорівнює 2 та більш застосовуваний варіант з F , що дорівнює 2 та з S , що дорівнює 2. Більший розмір агрегації є недоцільним [2, 3].

На рисунку 1.4 зображено приклад АШ з вхідним значенням 224 на 224 на 64, який агреговано з фільтром розміром, що дорівнює 2 та кроком, що дорівнює 2 та вихідним значенням 112 на 112 на 64 [2, 3].

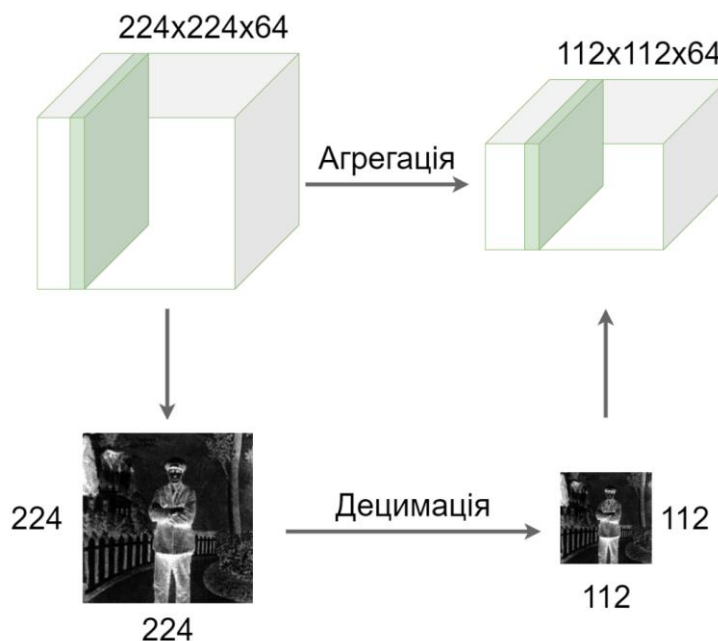


Рисунок 1.4 – Приклад АШ

Отже, АШ – це ефективний спосіб нелінійного децимації. Як і ЗШ він має рецептивне поле та крок згортки, але це не додає ніяких навчальних параметрів. АШ зазвичай ставиться після ЗШ. Рецептивне поле нейрона в ЗШ двовимірне. Кожен нейрон виводить максимум свого сприйнятливого поля. Зазвичай крок схожий на розмір сприйнятливого поля. Приймальні поля не перетинаються, а є дотичними. У більшості випадків крок та розмір рецептивного поля дорівнює 2 на 2. Операція максимізації підсилює найбільш наявну особливість (візерунок) свого рецептивного поля і видає решту. Після виявлення функції її грубе розташування відносно інших особливостей важливіше, ніж її точне місцезнаходження. ЗШ ефективно зменшує просторовий розмір представлення і не додає нових параметрів – зменшуючи їх для останніх шарів, роблячи обчислення більш ефективним [2, 3].

Через свою деструктивність – викидання 75% вхідної інформації у разі невеликого сприйнятливого поля 2 на 2 – поточна тенденція надає перевагу складеним ЗШ [4].

Нейрони в **повноз'єднаному шарі (ПШ)** мають повне з'єднання з усіма активаціями попереднього шару, як це спостерігається у звичайних нейронних мережах. Отже, їх активацію можна обчислити множенням матриць з подальшим вектором зміщення [2, 3].

Таким чином, в даному розділі було розглянуто будову типової звичайної нейронної мережі та типової згорткової нейронної мережі. Розглянуту схематичні зображення зазначених мереж. Також описано будову типової згорткової нейронної мережі з позначенням на схемі основних шарів (INPUT, CONV, RELU, POOL, FC). Був здійснений аналіз зазначених шарів та описано їх основні властивості.

2 ПОБУДОВА АЛЬТЕРНАТИВНОГО МЕТОДУ КОДУВАННЯ

2.1 Загальна архітектура альтернативної ЗНМ

За принципом нейронної мережі прямого поширення, ЗНМ складається з нейронів з ваговими значеннями та вектором зміщеннями. Нейрони ЗНМ все ще містять функцію активації і вся мережа виражає єдину диференційовану функцію оцінки. Положення пікселя має значення в порівнянні з багат шаровим перцептроном Румельхарта. Він отримує тривимірний вхід простору – значення Z -го каналу пікселя або виникнення Z -ї особливості ЗШ у відповідному положенні [2].

ЗШ та АШ локально з'єднані з виходами попереднього шару, розпізнаючи або збільшуючи локальні візерунки на зображенні. АШ зазвичай ставиться після ЗШ. Ця пара шарів багаторазово укладається один на одного завершуючи ПШ нагорі. Позиції нейронів у попередньому шарі завжди розглядаються як входи для останніх шарів. ПШ підключений до всіх виходів останнього АШ. Виходи останнього АШ вже повинні представляти складні структури та форми. Зазвичай ПШ завершується з одним або двома шарами в завершальному класі [2].

Типовою архітектурою для даного випадку може бути наступна послідовність шарів: вхідний шар (ВШ), ЗШ, АШ, ЗШ, АШ, ПШ, ПШ. Останні дослідження пропонують складати багато ЗШ разом із меншою кількістю АШ [2].

2.2 Зворотне поширення помилки

Вхідні дані або активації передаються наступним шарам, далі обчислюється скалярний добуток над яким застосовується функція активації. Можливо, може бути присутнім відбір проб мережі з використанням АШ. В кінці укладаються два

або три повністю з'єднаних шари. Для використання алгоритму градієнтного спуску має бути визначений градієнт [2].

Звичайний алгоритм зворотного розповсюдження застосовується з двома технічними оновленнями. Класичний алгоритм зворотного розповсюдження обчислював би різні часткові похідні вагових значень, що належать нейронам в одному фільтрі; однак вони повинні залишатися однаковими. Тому похідні функції втрат відносно вагових значень нейронів, що належать до однієї і тієї ж карти ознак, складаються разом [2].

Зворотне поширення помилки спрямовується лише до тих нейронів, які не були відфільтровані з операцією максимізації [2].

2.3 Автокодер

Автокодер – це нейронна мережа прямого поширення де вихідне значення мережі дорівнює вхідному, її мета реконструювати власні входи. Отже, автокодери належать до групи моделей навчання без нагляду [1]. Прийнято, що автокодери складаються з ВШ l_0 одного або багатьох прихованих шарів $l_1, \dots, l_{k-1}$ і вихідного шару (ВиШ). Так як ідея застосування автокодерів схожа з обмеженою машиною Больцмана (ОМБ) прийнято застосовувати правило $|l_1| = |l_{k-1}|$. Кожен автокодер складається з кодеру та декодеру [2].

Кодер може бути використаний для стиснення. На відміну від методу головних компонент (МГК) аналіз обмежено лінійним відображенням, де кодер являє собою базову нелінійну структуру даних [5]. Активації l_c шару можуть бути використані для подальшої класифікації. ПШ додаються з розміром останнього визначеного шару до відповідної кількості міток. Для подальшої роботи використовуються типові алгоритми навчання [2].

Поглиблений автокодер (ПА) складається з багатьох шарів, які розташовуються один за одним, що дозволяє визначати більш складні та нелінійні структури даних. Оскільки може бути досить важко налаштувати ПА мережі зазвичай використовують навчальну процедуру, яка складається з двох етапів. Перший етап визначає попереднє значення навчання, де кожний шар $l_1, \dots, l_c$ попередньо навчений. Як приклад використовується кодер l_1 та пара з 10 значень. Метою даного етапу є знаходження представлення l_0 в l_1 використовуючи необхідний оптимізатор. Ваги l_0 до l_1 та l_1 в l_0 можуть бути пов'язані враховуючи ОМБ. Коли в l_1 кодовано необхідне представлення входів l_0 пара l_1 та l_2 попередньо тренується поки не буде досягнуто пари l_{c-1}, l_c . Другий етап описує тонке налаштування, мережа повністю з'єднана та налаштована. У випадку класифікації системи, кодування входних відміток даних може бути використано для класифікації навчання або для всієї мережі [2].

Отже, в даному розділі було розглянуто будову альтернативної згорткової мережі для кодування та декодування мовного сигналу. Розглянуто загальну архітектуру альтернативної згорткової мережі, яка включає наступні шари: ВШ, ЗШ, АШ, ЗШ, АШ, ПШ, ПШ. Також розглянуто важливий елемент альтернативної згорткової мережі, як визначення зворотного визначення помилки. В кінці розділу описано загальний принцип роботи автокодера та його поглибленого режиму роботи.

З РЕАЛІЗАЦІЯ АЛЬТЕРНАТИВНОГО МЕТОДУ КОДУВАННЯ З МОДИФІКАЦІЄЮ

3.1 Частотно-часова згортка

Загальноприйнятим методом побудови ЗНМ для розпізнавання мови є використання аналізу частоти. Також досить часто застосовується спектральна згортка. Далі буде розглянуто альтернативний метод розпізнавання мови за допомогою ЗНМ. Даний метод був запропонований науковцями з Технологічного університету Ченстохова, Республіка Польща.

Особливістю даного методу є використання трьох ЗШ. Кожен з ЗШ відповідає за відповідну характеристику сигналу, а саме часову згортку, згортку спектральних коефіцієнтів (ЗСК) та спектральну згортку загалом. Використання часової згортки допомагає позбавитись проблем пов'язаних з часовими розбіжностями. Проблеми з спектральним зсувами, пов'язані з різною довжиною вокальних пачок, вирішуються за допомогою використання частотної згортки. Використання згортки спектральних коефіцієнтів зменшує чутливість до сторонніх завад та шумів, а також дозволяє адаптувати частотний аналіз мовного сигналу, який близький до сприйняття сигналу людським вухом. Спектр аналізує сигнал лінійно, тоді як ЗСК виконує лінійний аналіз до частоти 1 кГц, а на 1 кГц і вище характеристика стає експоненціальною. На рисунку 3.1 схематично зображено будова зазначеного методу [2].

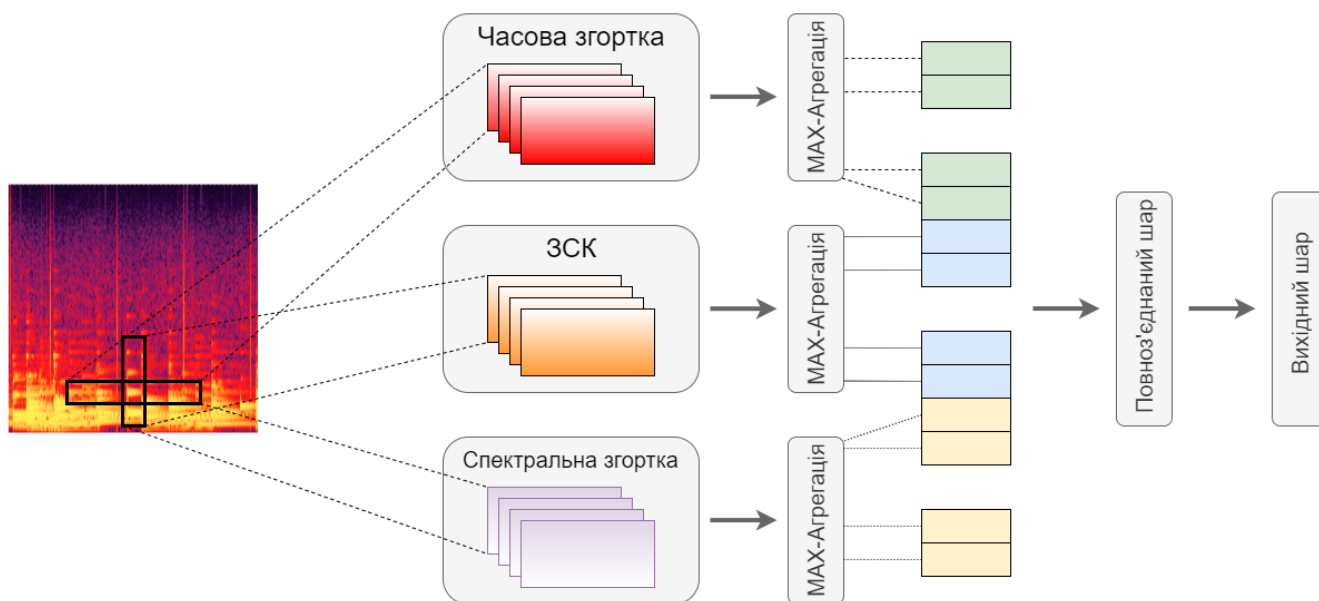


Рисунок 3.1 – Альтернативний метод кодування мовного сигналу

Даний метод дозволяє кодування мови за допомогою зображення RGB формату. Для кодування ЗСК було виділено R компонент, для кодування часових характеристик застосовано G компонент, і для кодування загального спектру G компонент відповідно. Згортка всіх трьох компонентів R, G, B відбувається паралельно. Карта активації, яка використовується для кодування, має різні розміри. При створенні кольорового зображення важливою частиною є масштабування RGB компонентів до єдиного розміру. Отже, було визначено розмір для окремих звуків 120 на 120 пікселів. Як приклад, групою науковців було взято слово “seven”. Його часову характеристику можна побачити на рисунку 3.2.[2].



Рисунок 3.2 – Часова характеристика мовного сигналу [2]

Також на рисунку 3.3 можна побачити значення ЗСК після нормалізації. На рисунку 3.4 зображено двовимірну часову діаграму після нормалізації. На рисунку 3.5 визначено спектральну характеристику після нормалізації [2].

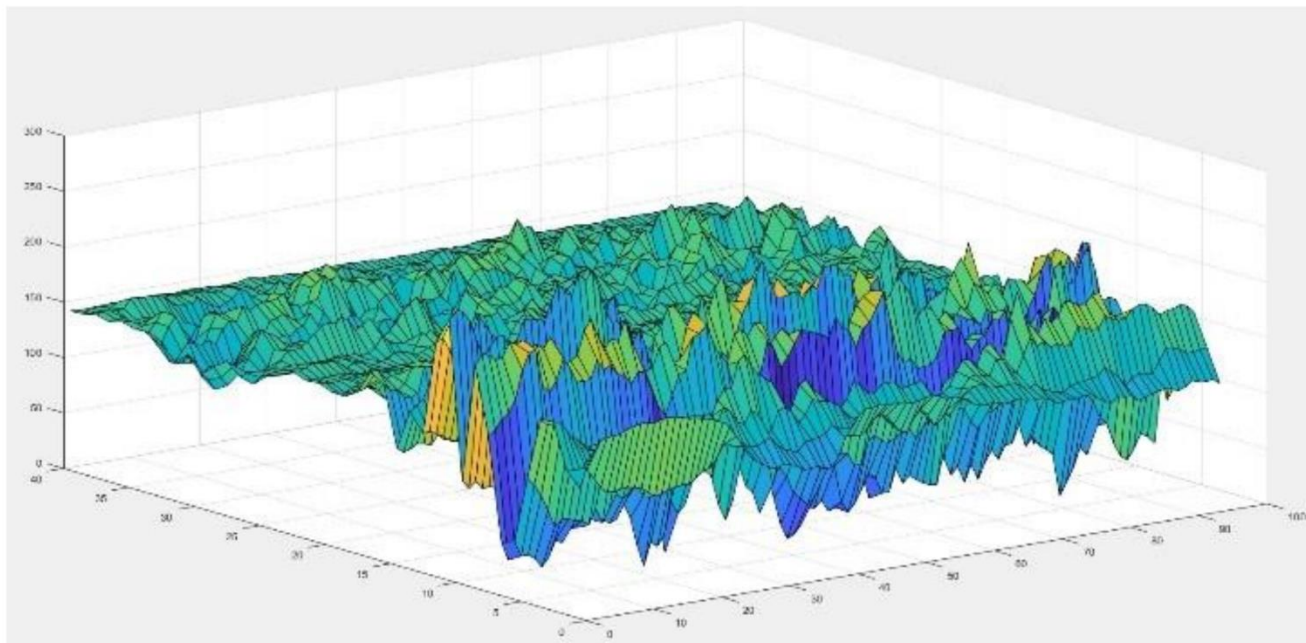


Рисунок 3.3 – Значення ЗСК після нормалізації [2]

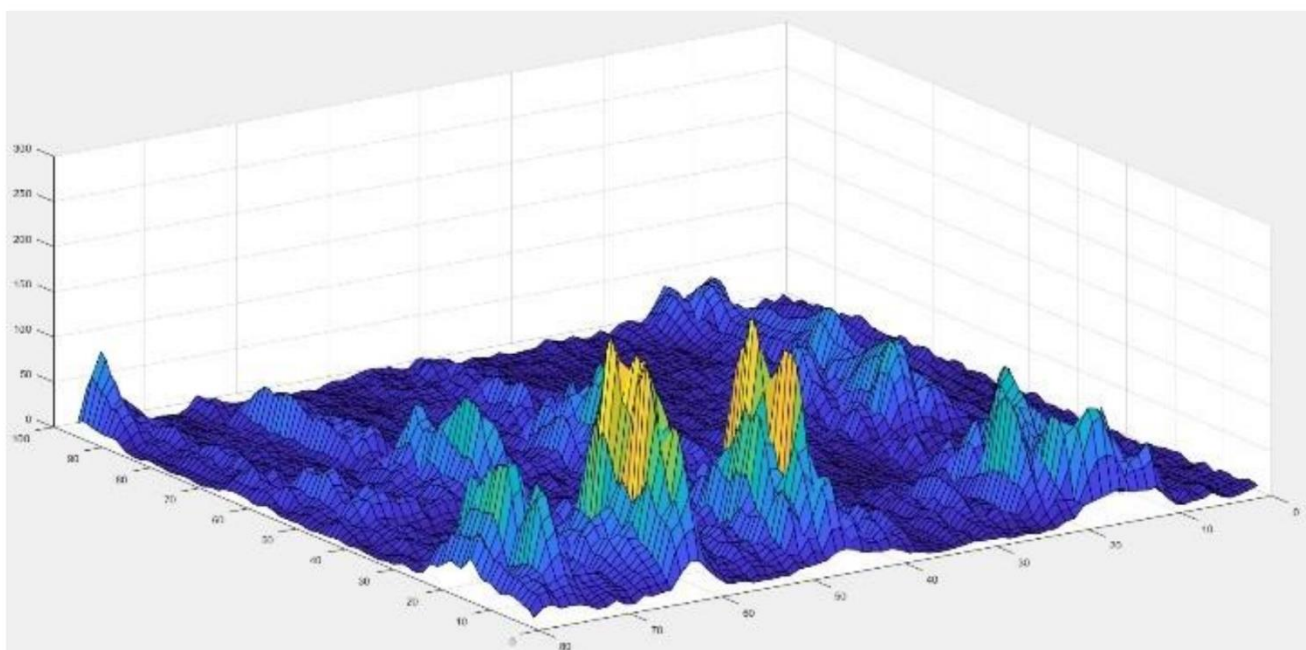


Рисунок 3.4 – Двовимірну часову діаграму після нормалізації [2]

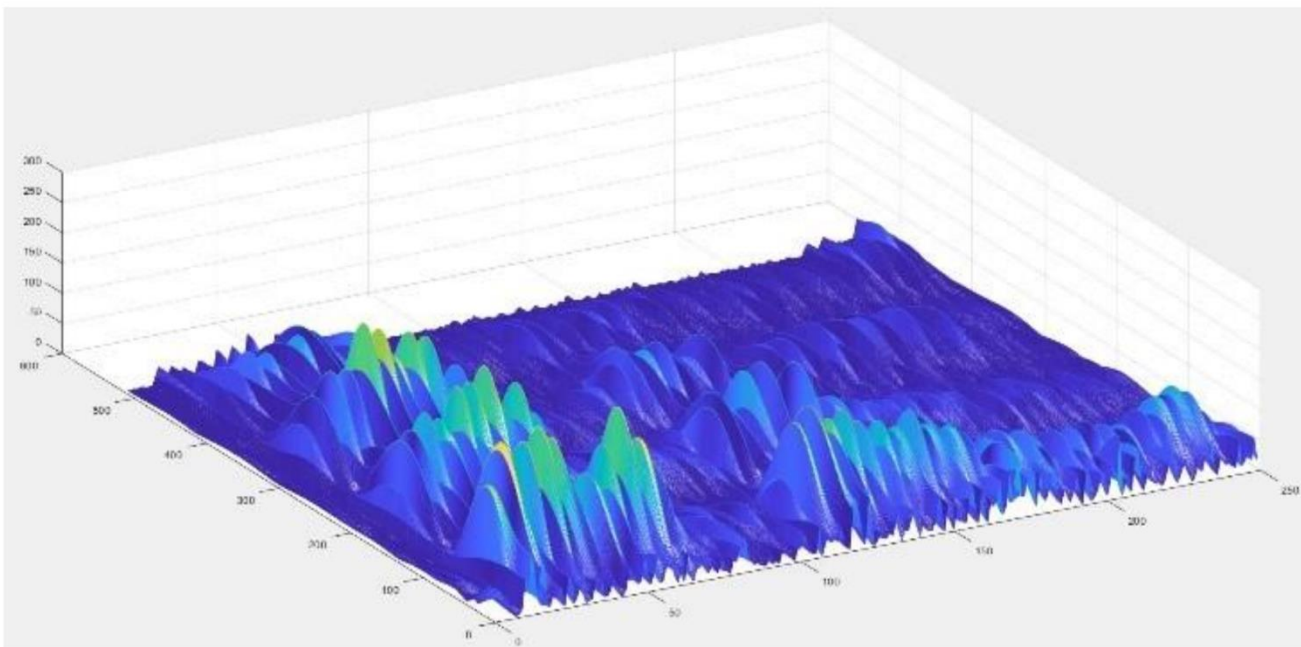


Рисунок 3.5 – Спектральну характеристику після нормалізації [2]

Для кожного компоненту було визначено RGB зображення, які є візуальним представленням фонemi. Також було визначено, що кожен склад слова має бути кодований окремо [2].

3.2 Структура альтернативної ЗНМ

Запропонована ЗНМ складається з 15 шарів. Перший ЗШ включає в себе 64 фільтри розміром 9 на 9. Три ЗШ відповідають за кодування інформації, яка передається до ПШ. Дана схема зображена на рисунку 3.6 [2].

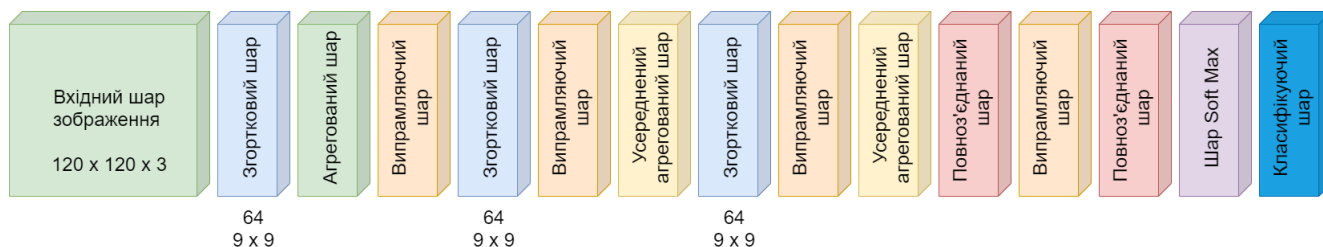


Рисунок 3.6 – Структура альтернативної ЗНМ [2]

Останній ПШ та класифікуючий шар (КШ) мають X можливих вихідних варіантів [2].

Дані X варіанти залежать від навчального набору даних ЗНМ. Вхідні дані можуть бути як цілі слова, так і склади. Для окремих слів X , це кількість визначених слів. Для складів X , це кількість різних складів визначених з навчального набору даних. Бувають випадки коли алгоритм розділення комбінує окремі звуки з тишею [2].

3.3 Результати дослідження альтернативного методу кодування

Дослідження включали два важливі питання. Перший стосувався правильності роботи алгоритму, реалізуючи поділ мови на склади. Був застосований метод, який враховує енергію сигналу та частоту сигналу для розглянутого нерухомого фрагмента. Розділення на окремі склади проводили для експериментально вибраних порогових значень. Для кожного ізольованого складу (який також міг включати тишу) зображення створювалося як графічний візерунок. Експеримент проводили для ізольованих слів і безперервного мовлення. Для безперервного мовлення прийнято різне число слів. Таблиця 3.1 показує результати експерименту щодо поділу на склади. Дослідження проводилось на базі бази даних університету та публічних аудіокниг [2].

Таблиця 3.1 [2]

Номер експерименту	Тип експерименту	Кількість слів (1) та речень (2 - 4)	Правильна вибірка, %
1	Окремі слова	70	98.1
2	безперервна мова (від 3 до 7 слів)	65	86.7
3	безперервна мова (від 6 до 12 слів)	65	80.5
4	безперервна мова (від 10 до 20 слів)	65	69.7

Найбільш поширені в літературі твори, які використовують для аналізу часові та частотні характеристики. Дане дослідження включає новий компонент в аналізі сигналу ЗСК. Перевагою цієї методики є аналіз звукового аналізу так само, як це робить людське вухо [2].

Отриманий результат важко порівняти з іншими існуючими методами, оскільки попередні підходи включали лише одну або дві складові (частотну та спектральну). Крім того, ці методи використовувались для зображень у масштабах сірого, в той час як даний метод використовує потрійний аналіз (крім ЗСК) для RGB зображень [2].

Зазначається, що найкращі результати були отримані для відокремлених слів – 98%. Очевидно, що результати, отримані для безперервного мовлення, мають менший показник розпізнавання. Проте результат 70 – 80% задовільний [2].

Другий тип досліджень стосувався оцінки ефективності розпізнавання мовлення для ізолюваних слів та для постійного мовлення. Записані дані були поділені на навчальні дані та тестові дані у співвідношенні 70 до 30. Також було вивчено час навчання нейронної мережі та кількість часових проходів, необхідних для правильного навчання. Результати основного експерименту наведені в таблиці 3.2 [2].

Таблиця 3.2 [2]

Номер експерименту	Тип експерименту	Час навчання, с	Кількість часових проходів, %	Відношення неправильних слів, %
1	Окремі слова	1920	125	4.2
2	безперервна мова (від 3 до 7 слів)	2658	317	9.7
3	безперервна мова (від 6 до 12 слів)	8280	428	11.3
4	безперервна мова (від 10 до 20 слів)	10.065	641	14.8

3.4 Модифікація альтернативного методу кодування мови

Кодування мови було і залишається головною проблемою в області цифрової обробки мови. Кодування мови – це процес перетворення мовного сигналу у більш компактну форму, яка потім може передаватися зі значно меншою пам'яттю. Мотивацією цього є той факт, що доступ до необмеженої кількості пропускної здатності неможливий. Тому виникає необхідність кодувати та стискати мовленнєві сигнали. Стиснення мови потрібно для міжміського зв'язку, якісного зберігання мови та шифрування повідомлень. Наприклад, у цифрових стільникових технологіях багатьом користувачам потрібно ділитися однаковою смугою частот. Використання стиснення мови дозволяє більшій частині користувачів поділитися доступною системою. Інший приклад, коли потрібно стиснути мовлення, - це цифрове зберігання голосу [6].

Кодування мови – це збитковий тип кодування, що означає, що вихідний сигнал не точно звучить як вхідний. Мовлення виробляється при опущенні велуму, щоб акустично поєднуватися з голосовим трактом. Мовленнєві сигнали складаються з декількох послідовностей звуків. Кожен звук можна розглядати як унікальну інформацію. Існують голосні та приголосні типи мовних звуків. Принципова відмінність цих двох типів мовленнєвих звуків виходить від способу їх продукування. Коливання голосових зв'язок видають голосні звуки. Швидкість, з якою вібрують голосові зв'язки, диктує висоту звуку. З іншого боку, приголосні звуки не покладаються на вібрацію голосових зв'язок. Приголосні звуки створюються звуженням голосового тракту. Голосові зв'язки залишаються відкритими, а звуження голосового тракту виштовхує повітря, щоб видавати незвукові звуки [6].

Для аналізу та синтезу мовних сигналів буде використовуватися техніка LPC. Цей метод використовується для успішної оцінки основних мовних

параметрів, таких як тон голосу, форманти та спектральні характеристики. Структурну схему вокодера можна побачити на рисунку 3.7 [7].

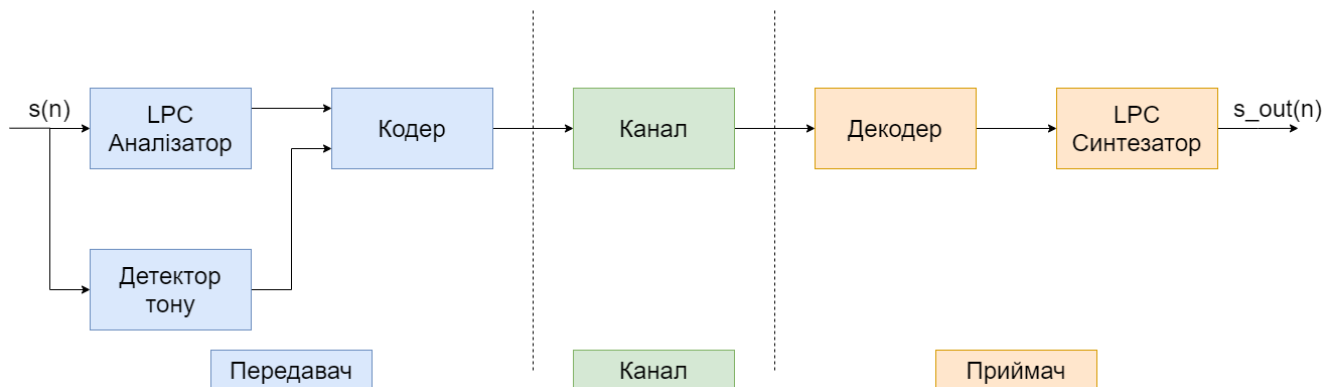


Рисунок 3.7- Структурну схему вокодера

Принцип використання LPC полягає у зведенні до мінімуму суми різниць у квадраті між початковим мовним сигналом та передбачуваним мовним сигналом протягом обмеженої тривалості. Це може бути використане для отримання унікального набору коефіцієнтів предиктора [7]. Ці коефіцієнти предиктора, як правило, оцінюються в кожному кадрі, який зазвичай становить 20 мс [7].

Запропонований метод кодування мовного сигналу застосовуючи потрібну згортку трьох основних характеристик сигналу показав свою ефективність в лабораторних умовах з можливістю подальшого використання в промисловості. Отже, враховуючи методологію зазначеного методу та позитивні аспекти проведених досліджень можна використати даний підхід для кодування мовних сигналів в каналах зв'язку. Для прикладу обрано базовий вокодер LPC-10, який є основою для багатьох наступних вокодрів та досить простий в застосуванні. На рисунку 3.8 зображено бітовий потік мовного сигналу після кодера [7].

На момент написання даної роботи, здійснюється визначення кількісних та якісних характеристик відповідно до запропонованої модифікації. Передбачається, що даний метод на основі позитивних результатів попереднього методу дозволить підвищити якість кодування та декодування мови в умовах апіорної невизначеності або незадовільних умовах розповсюдження сигналу при передачі його в каналах зв'язку.

Таким чином, в даному розділі було розглянуто процес частотно-часової згортки для запропонованого альтернативного методу кодування та декодування мовного сигналу. Також проаналізовано структуру згорткової нейронної мережі для запропонованого методу. Було проаналізовано результати запропонованого методу, які показали позитивні показники зазначені даному розділі. В кінці розділу було запропоновано модифікацію альтернативного методу згорткової нейронної мережі для кодування та декодування мовного сигналу в каналах зв'язку на прикладі вокодеру LPC.

Висновки

Таким чином, в першому розділі було розглянуто будову типової звичайної нейронної мережі та типової згорткової нейронної мережі. Розглянуту схематичні зображення зазначених мереж. Також описано будову типової згорткової нейронної мережі з позначенням на схемі основних шарів (INPUT, CONV, RELU, POOL, FC). Був здійснений аналіз зазначених шарів та описано їх основні властивості.

В другому розділі було розглянуто будову альтернативної згорткової мережі для кодування та декодування мовного сигналу. Розглянуто загальну архітектуру альтернативної згорткової мережі, яка включає наступні шари: ВШ, ЗШ, АШ, ЗШ, АШ, ПШ, ПШ. Також розглянуто важливий елемент альтернативної згорткової мережі, як визначення зворотного визначення помилки. В кінці розділу описано загальний принцип роботи автокодеру та його поглибленого режиму роботи.

В третьому розділі було розглянуто процес частотно-часової згортки для запропонованого альтернативного методу кодування та декодування мовного сигналу. Також проаналізовано структуру згорткової нейронної мережі для запропонованого методу. Було проаналізовано результати запропонованого методу, які показали позитивні показники зазначені даному розділі. В кінці розділу було запропоновано модифікацію альтернативного методу згорткової нейронної мережі для кодування та декодування мовного сигналу в каналах зв'язку на прикладі вокодеру LPC.

Список літератури

1. Rawat W., Wang Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. – [Електронний ресурс]. – Режим доступу: https://www.mitpressjournals.org/doi/pdfplus/10.1162/neco_a_00990 (дата звернення 20.03.2020 р.).
2. Kubanek M, Janusz Bobulski J, Kulawik J A Method of Speech Coding for Speech Recognition Using a Convolutional Neural Network. – Режим доступу: <https://www.mdpi.com/2073-8994/11/9/1185/htm#B18-symmetry-11-01185> (дата звернення 22.03.2020 р.).
3. Convolutional Neural Networks (CNNs/ConvNets). – [Електронний ресурс]. – Режим доступу: <https://cs231n.github.io/convolutional-networks/> (дата звернення 27.03.2020 р.).
4. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. – [Електронний ресурс]. – Режим доступу: <https://arxiv.org/pdf/1512.03385.pdf> (дата звернення 28.03.2020 р.).
5. Hinton G.E.; Salakhutdinov R.R. Reducing the dimensionality of data with neural networks – [Електронний ресурс]. – Режим доступу: <https://www.cs.toronto.edu/~hinton/science.pdf> (дата звернення 20.03.2020 р.).
6. Alwan A., Ozgu Ozun O., Steurer P., Thell D., Wideband Speech Coding With Linear Predictive Coding (LPC) – [Електронний ресурс]. – Режим доступу: <http://www.seas.ucla.edu/spapl/projects/ee214aW2002/1/report.html> (дата звернення 05.04.2020 р.).
7. Rabiner L., Schafer R. Digital Processing of Speech Signals – [Електронний ресурс]. – Режим доступу: <https://asa.scitation.org/doi/pdf/10.1121/1.384160> (дата звернення 20.04.2020 р.).