



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ  
ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра Экспериментальной физики атмосферы

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**  
**(бакалаврская работа)**

На тему «Прогноз осадков по данным ДМРЛ с использованием алгоритмов машинного обучения»

Исполнитель                   Фурлетов Никита Максимович  
(фамилия, имя, отчество)

Руководитель                доктор физико-математических наук, профессор  
(ученая степень, ученое звание)

Кузнецов Анатолий Дмитриевич  
(фамилия, имя, отчество)

«К защите допускаю»

Заведующий кафедрой

(подпись)

доктор физико-математических наук, профессор  
(ученая степень, ученое звание)

Кузнецов Анатолий Дмитриевич  
(фамилия, имя, отчество)

«   » мая 2022 г.

Санкт-Петербург  
2022

## Оглавление

ВВЕДЕНИЕ.....	4
ГЛАВА 1. РАДИОЛОКАЦИОННЫЙ МЕТОД ИЗУЧЕНИЯ АТМОСФЕРЫ.....	8
1.1 Базовые определения работы доплеровского метеорологического локатора. ....	8
1.2 Принцип работы доплеровского метеорологического локатора. ....	9
1.3. Радиолокационная отражаемость. ....	10
1.4 Доплеровская радиальная скорость.....	13
1.5. Радиолокационное измерение осадков. ....	17
ГЛАВА 2. Машинное обучение в задачах метеорологии.....	20
2.1. Сверточная нейронная сеть.....	21
2.2. U-NET сверточная нейронная сеть. ....	40
ГЛАВА 3. Построение и использование U-NET сверточной нейронной сети для краткосрочного прогнозирования передвижения облачности по данным доплеровского метеорологического локатора.....	46
3.1. Постановка задачи.....	46
3.2. Обзор и обработка данных.....	47
3.3. Выбор модели, её настройка и тестирование. ....	50
3.4. Визуализация результатов. ....	52
ЗАКЛЮЧЕНИЕ.....	54
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ .....	55

## ВВЕДЕНИЕ

ДМРЛ. Известным фактом является то, что метеорологический радиолокатор предоставляет данные о атмосфере в радиусе от 125 до 300 км в режиме реального времени. Такая метеорологическая информация является наиболее интересной для изучения погодных процессов.

Есть несколько особенностей данных МРЛ. Во-первых, время измерения на локах в Российской Федерации составляет 10 минут, что позволяет следить за малейшими изменениями облачных систем. Во-вторых, радиолокатор позволяет узнать информацию об облаке на уровнях до 15 км. Изучив данные, можно определить характер и фазу осадков, вектор движения облачной массы, турбулентность и так далее. В-третьих, пространственное разрешение составляет от 0.5 до 1 км. Данные ДМРЛ предоставляют точную и полную информацию об атмосфере в большом радиусе от лока.

Современные метеорологические локаторы решают следующие важные практические задачи:

- Идентификация метеоявлений, связанных с облачностью и осадками, информационное обеспечение сверхкраткосрочного прогноза погоды и штормовых предупреждений, предупреждений по аэродрому, предупреждений о сдвиге ветра.
- Измерение характеристик осадков радиолокационными методами и использование этих данных в гидрологических расчётах и прогнозах,
- Обеспечение радиолокационной информацией численных моделей прогноза погоды для их инициализации и верификации.

[1]

Машинное обучение. Для решения задач прогнозирования всё больше используют новый инструмент – машинное обучение.

Алгоритмы машинного обучения позволяют решать разнообразные задачи: от подбора фильма для просмотра до просчёта лучшего хода в шахматной задаче.

В машинном обучении можно выделить три вида: обучение с учителем, обучение без учителя и глубокое обучение. В первом случае, на вход модели подаётся начальная информация и цель анализа. К примеру, у нас есть информация об изменении метеорологических параметров в определённый промежуток времени, а нашей целью является прогноз температуры на следующие три дня. Мы задаём нашей модели имеющиеся данные, и те значения температуры, которые были по факту – что является целью нашего анализа. Этот метод решения называют регрессией. Другим примером в первом случае, будет прогнозирование определённого явления, по тем же имеющимся данным. В таком случае, в цель модели, мы будем подавать 1 и 0. Единица будет означать, что явление будет, а ноль – его отсутствие. Такой пример задачи называют бинарной классификацией. Если же классов несколько, то задача будет многоклассовой классификацией.

Второй вид обучения – обучение без учителя. Обучение строится на том, что программе не известна цель обучения, а имеется только определённый массив данных. В таком случае, алгоритм сам находит взаимосвязи в данных. Используя модели кластеризации, в данных можно выделить дни с определённой синоптической ситуацией, к примеру дни с прохождением фронтальных систем или ясные дни в центре антициклона. Так же для более удобной работы и визуализации, подходят методы понижения размерности. В таком случае из десятков метеорологических характеристик, получаем несколько информативных признаков, которые будут более удобны для визуализации и обработки моделями в дальнейшем.

Последний вид обучения – глубокое обучение. Используется для работы с большими объёмами данных. В основе данного вида, лежат искусственные нейронные сети. Такие сети решают большое количество задач в различных сферах, например, распознавание речи, машинный

перевод, компьютерное зрение. Базовая часть сети – нейрон. Математически нейрон представляют как линейную комбинацию всех входных сигналов – функция активации. Несколько нейронов можно сложить в один слой, в таком случае, получится однослойная нейронная сеть. Сложив несколько слоёв, получаем многослойную нейронную сеть. Именно такие сложные сети используются при решении сложных задач, так как обладают большими возможностями при нахождении взаимосвязей в данных.

Актуальность работы. В метеорологии, задача краткосрочного прогноза для авиации является одной из самых востребованных задач. Использование алгоритмов машинного обучения способны улучшить точность прогнозирования перемещения воздушных масс на основе данных метеорадиолокатора.

Объект исследования – данные доплеровского метеорологического локатора

Предмет исследования – нейронные сети в задачах прогнозирования перемещения облачности

Цель – использовать нейронные сети для прогнозирования перемещения облачности

Для достижения поставленной цели необходимо решить следующие задачи:

1. Обработать и изучить имеющиеся данные.
2. Поставить задачи, которые способны решить алгоритмы машинного обучения
3. Решить задачи, используя нейронные сети.
4. Проанализировать результат и сделать вывод о применении алгоритмов машинного обучения в задачах прогнозирования передвижения облачности.

База исследования. Данные ДМРЛ получены из ООО “Институт радарной метеорологии”, ИРАМ.

Структура работы: введение, три главы, заключение, список использованной литературы и приложения.

Во введение работы обоснована актуальность выбранной темы исследования, определена цель и задачи исследования.

В первой главе работы «Радиолокационный метод изучения атмосферы»: дан развёрнутый ответ на вопрос о том, что является метеорологической радиолокационной информацией.

Во второй главе работы: «Машинное обучение в задачах метеорологии»: представлены нейронные сети, использовавшиеся для работы и их применение на основе данных метеорологического радиолокатора.

Во второй главе работы: «строение и использование U-NET сверточной нейронной сети для краткосрочного прогнозирования передвижения облачности по данным доплеровского метеорологического локатора»: были проведены: постановка задачи для нейронной сети U-NET, обзор данных и их обработка, настройка, обучение модели, а также её тестирование и визуализация результатов.

В заключение работы сделаны основополагающие выводы по результатам исследования.

## ГЛАВА 1. РАДИОЛОКАЦИОННЫЙ МЕТОД ИЗУЧЕНИЯ АТМОСФЕРЫ

### 1.1 Базовые определения работы доплеровского метеорологического локатора.

В течение многих лет радар использовался для получения информации об интенсивности, количестве и фазовом составе выпадающих осадков. Это применение радара хорошо известно из многочисленных метеорологических сайтов и приложений. Значительное улучшение возможностей метеорологического радара произошло, когда инженеры разработали новые радары, которые могли измерять доплеровский сдвиг частоты в дополнение к величине радиолокационной отражаемости. Доплеровский сдвиг частоты важен, поскольку он связан с радиальной скоростью осадков, переносимых ветром (составляющая ветра, движущаяся либо в сторону, либо в сторону от радиолокационной установки). Поскольку торнадо, мезоциклоны, ураганы и другие опасные погодные явления имеют тенденцию к вращению, измерение радиальной скорости ветра, как функции угла обзора позволит определить вращающиеся погодные условия. Вращение указывается, когда измерение доплеровского сдвига частоты показывает, что ветер приближается к радару под одним углом и удаляется от него под другим углом.

Импульсные доплеровские метеорологические радары, используемые на сети Росгидромета, производят количественные измерения осадков, предупреждают о потенциальных наводнениях или опасном граде, определяют скорость и направление ветра, указывают на наличие фронтов сдвига ветра и порывов, отслеживают штормы и предоставляют другую метеорологическую информацию. В дополнение к измерению осадков (по интенсивности эхо-сигнала) и радиальной скорости (по доплеровскому смещению частоты) ДМРЛ-С на сети Росгидромета также измеряют разброс радиальной скорости (разница между максимальной и минимальной

скоростями) частиц осадков в пределах каждого радара и в каждой ячейке. Разброс радиальной скорости указывает на турбулентность ветра.

На текущий момент в составе сети обеспечивается сбор, обработка и архивация первичных данных 42 радиолокаторов ДМРЛ-С Росгидромета, которые поступают, обрабатываются и архивируются в Центре сбора и обработки радиолокационной информации (ЦСО РЛИ) НТЦР ДМРЛ ФГБУ «ЦАО», включая новые ДМРЛ «Тюмень», «Красный Кут», «Псков», «Пермь», «Челябинск», а также ДМРЛ WRM200 «Ахун» (Сочи) и 28 зарубежных ДМРЛ. [2]

## 1.2 Принцип работы доплеровского метеорологического локатора.

Доплеровский метеорологический радиолокатор с двойной поляризацией ДМРЛ-С предназначен для обеспечения метеорологической информацией об облачности, осадках и связанных с ними явлениях погоды прогностических органов Росгидромета, АМЦ, АМСГ и Центров УВД гражданской авиации, а также других потребителей радиолокационной метеорологической информации.

Радиолокационные наблюдения на сети ДМРЛ-С проводятся синхронно по единому регламенту на всех радиолокаторах сети ДМРЛ Росгидромета. Каждый десятиминутный цикл включает производство наблюдений в 2-х режимах: в режиме «Отражаемость» и в режиме «Скорость», в каждом режиме данные наблюдений записываются в т. н. «объёмный файл». Радиус зоны радиолокационного обзора в режиме «Отражаемость» составляет 250 км. Радиолокационные наблюдения в режиме «Скорость» с высокой частотой повторения зондирующих импульсов ( $PRF \sim 1$  кГц) обеспечивают более точное измерение доплеровских характеристик (радиальной скорости  $V$ , ширины спектра  $W$ ), однако при этом сокращается дальность зондирования с 250 до 125 км.

В каждом обзоре радиолокатор последовательно проводит азимутальное круговое сканирование атмосферы на нескольких углах места антенны в диапазоне от 0 до 90°. В результате, в каждом режиме наблюдений р/л данные записываются в выходном «объёмном файле», в котором в сферической СК последовательно, для каждого элемента конического сечения, сохраняются координаты – дальность, азимут и угол места, а также измеренные р/л характеристики радиоэха. [1]

### 1.3. Радиолокационная отражаемость.

Изображения отражательной способности точно так же, как они звучат, поскольку они рисуют картину погоды из энергии, отраженной обратно к радару. Изображения отражательной способности составляют подавляющее большинство радиолокационных изображений.

Базовая отражательная способность, полученная при сканировании с наименьшим углом ( $1/2^\circ$ ), отлично подходит для съемки области вокруг радара в поисках осадков. Ниже представлено изображение базовой отражательной способности, показывающее осадки.

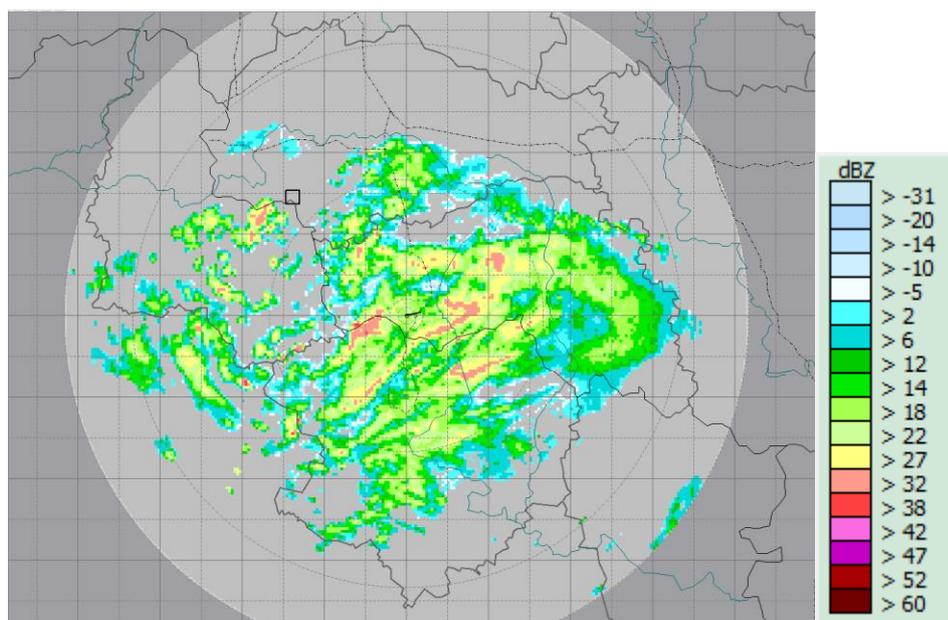


Рисунок 1.1. Радиолокационная отражаемость на уровне 1 км. ДМРЛ-С  
Белгород, 6:50 UTC 8 мая 2020 года.

Это изображение (выше) является образцом базового изображения отражательной способности с доплеровского радара в Белгороде, Белгородская область. Радар расположен в центре изображения.

Цвета представляют силу возвращенной энергии радара, выраженную в децибелетах (dBZ). Цветовая шкала расположена в правом нижнем углу каждого изображения. По мере увеличения значений dBZ увеличивается и интенсивность осадков.

Значение 20 dBZ обычно соответствует точке, в которой начинается небольшой дождь. Значения от 60 до 65 dBZ соответствуют уровню, при котором может выпасть град диаметром 2,5 см. Однако значение от 60 до 65 dBZ не означает, что в этом месте наблюдаются суровые погодные условия.

Суровые погодные условия могут возникать со значениями больше 60–65 дБЗ:

- Град полностью замерзший (без тонкого слоя воды на поверхности). «Сухой град» очень плохо отражает энергию и может привести к недооценке силы грозы.
- В определённых атмосферных явлениях возникают сильные потоки воздуха. Когда такие потоки образуются, то луч радара преломляется в землю (указывая на более сильные штормы, чем те, которые происходят на самом деле). Однако в худшем случае возникает субрефракция и луч выходит за пределы наиболее интенсивных областей штормов (указывая на более слабые штормы, чем те, которые происходят на самом деле).
- Доплеровские радары, не прошедшие калибровку. Радар может указывать на более сильные штормы, чем те, которые происходят на самом деле или наоборот, указывать на более слабые штормы, чем те, которые происходят на самом деле.

- Луч радара распространяется с расстоянием, что означает, что наиболее интенсивная часть отраженных от шторма отражений будет усреднена с более слабыми частями, что приведет к общему виду меньшей интенсивности.
- И, наконец, что не менее важно, луч радара увеличивается по высоте по мере увеличения расстояния от радара.

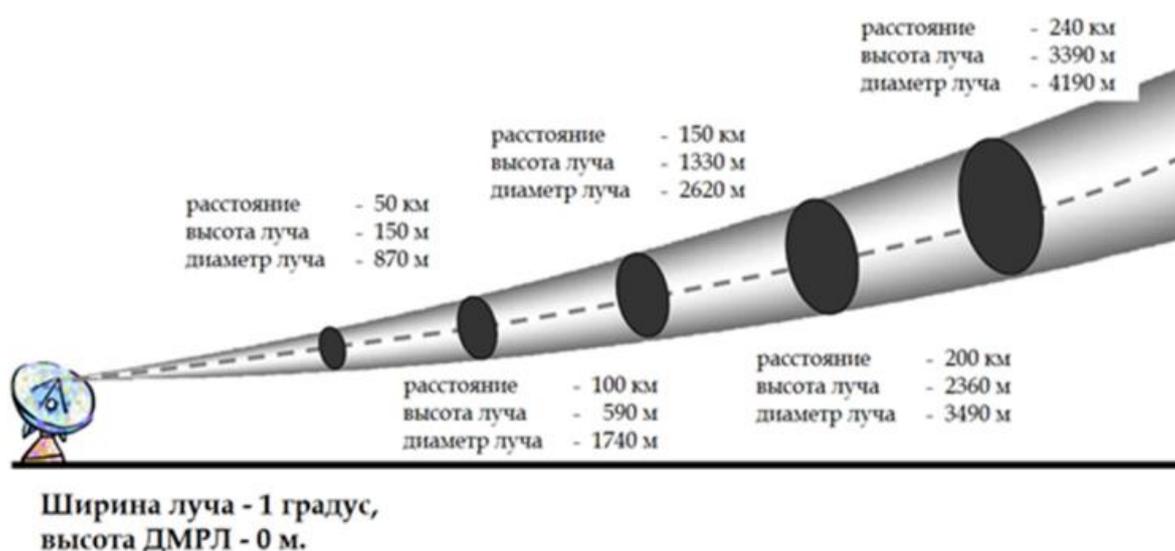


Рисунок 1.2. Подъем и расширение р/л луча с расстоянием от ДМРЛ

При увеличении расстояния радар видит все выше и выше во время шторма, и луч может промахнуться, и не задеть наиболее интенсивные участки.

Когда все возвратные сигналы со всех сканов высот скомпилированы, создается изображение, которое принимает самое высокое значение dBZ из всех высот, называемое комбинированной отражательной способностью. Это картина сильнейших возвратов со всех высот.

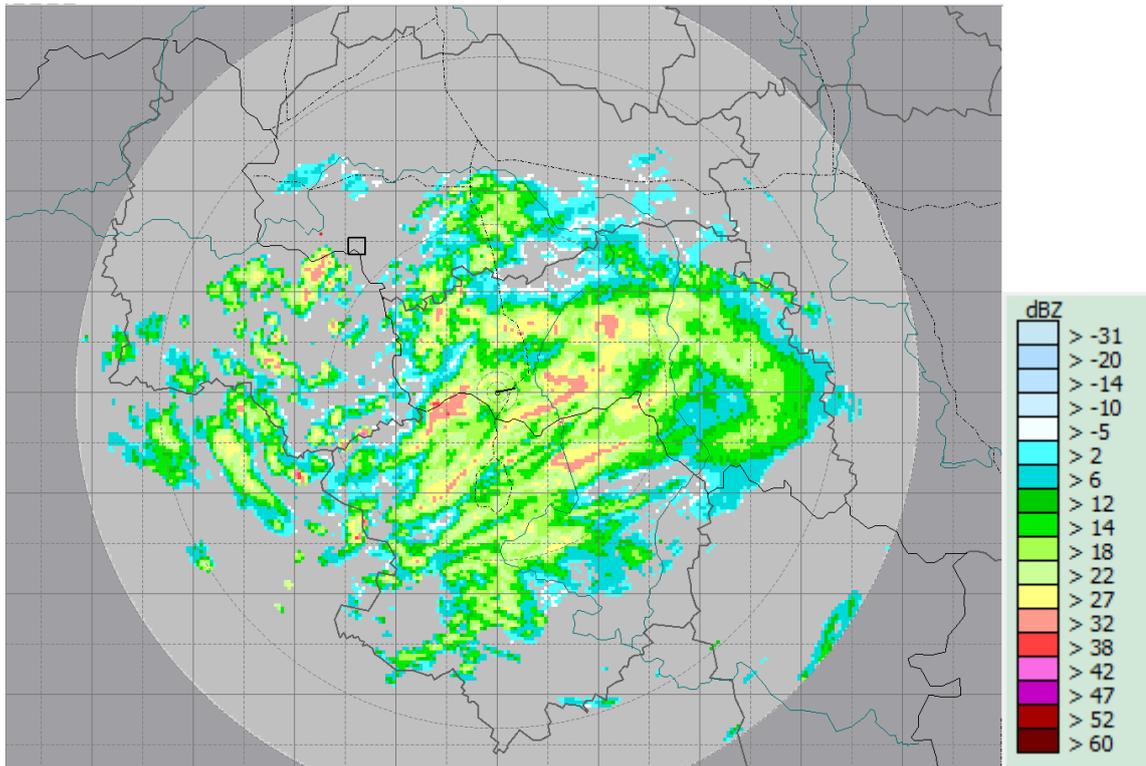


Рисунок 1.3. Максимальная радиолокационная отражаемость на уровне выше 1 км. ДМРЛ-С Белгород, 6:50 UTC 8 мая 2020 года.

По сравнению с базовой отражательной способностью композитная отражательная способность может выявить важные особенности структуры шторма и тенденции интенсивности шторма. Это важно, потому что часто во время развития сильных и сильных гроз в результате сильных восходящих потоков образуются участки, свободные от дождя (или участки с небольшим дождем).

#### 1.4 Допплеровская радиальная скорость.

Что отличало доплеровский радар от радаров предыдущего поколения, так это его способность обнаруживать движение. Движение, которое он видит, — это в первую очередь капли дождя, переносимые ветром, но он также может обнаруживать движения насекомых, птиц и частиц дыма.

Однако единственное движение, которое он может «видеть», называется радиальной скоростью. Это движение НЕ направление ветра, а часть движения ветра, которая движется либо прямо к радару, либо от него.

Движение ветра относительно радара разбивается на две составляющие: движение перпендикулярно лучу радара и движение по этому лучу (прямо к радару или от него).

На графике ниже ветер движется с юга на север (обозначено черными стрелками). К северу от радара (красная заливка) ветер удаляется от радара. По мере того, как радар перемещается из положения 1 в положение 2, луч становится все более и более совпадающим с общим потоком ветра.

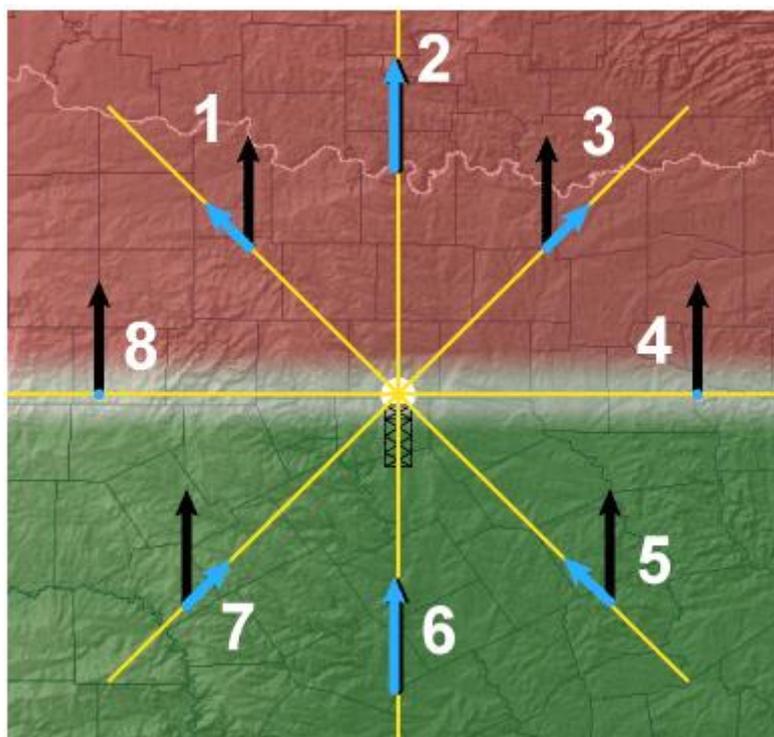


Рисунок 4. Изображение радиальной скорости ДМРЛ.

При этом радар «видит» увеличение радиальной скорости вдали от радара. В положении 2 радиальная скорость равна общей скорости ветра. Затем, когда радар перемещается в положение 3, радиальная скорость начинает уменьшаться. В положении 4 (и 8) ветер дует перпендикулярно лучу радара. Поскольку движения к радару (или от него) нет, он «видит» нулевое движение. Однако ветер в этих точках НЕ УСТОЙЧИВ, так как он все еще дует с юга. Это как раз область нулевой лучевой скорости. Позиции 5, 6 и 7, выделенные зеленым цветом, аналогичны позициям 1, 2 и 3,

отмеченным красным цветом, за исключением того, что ветер движется в сторону радара. Наибольшая радиальная скорость наблюдается в позиции 6, где ветер дует прямо на радар.

Эти наблюдаемые радиальные движения являются векторами, а это означает, что длина стрелок указывает на скорость ветра; чем длиннее стрелка, тем выше скорость. Доплеровский радар вычисляет базовую скорость на основе длины этих векторов и создает графическое отображение с цветовой кодировкой для отображения. На этих графиках скоростей красный цвет указывает на ветер, удаляющийся от радара, а зеленый цвет указывает на ветер, движущийся к радару.

Чем ярче красный и зеленый, тем больше радиальная скорость и больше репрезентация истинной скорости ветра. Стрелками на рисунках 1.5, 1.6 и 1.7 указаны примерные направления движения гидрометеоров на данных уровнях.

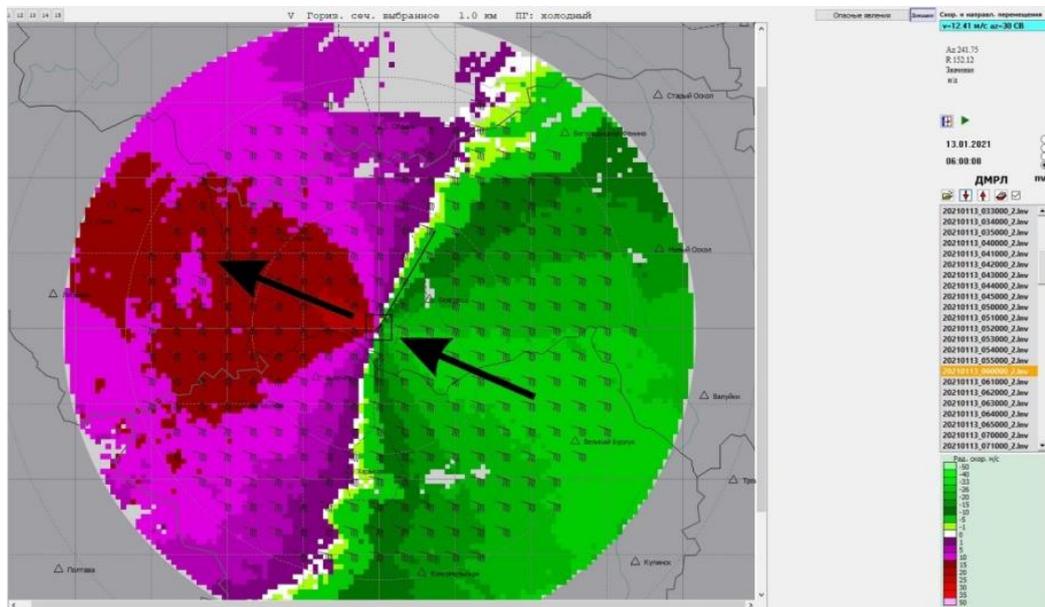


Рисунок 1.5. Карта радиальной доплеровской скорости на высоте 1 км.

ДМРЛ-С Белгород, 6:00 UTC 13 января 2021 г.

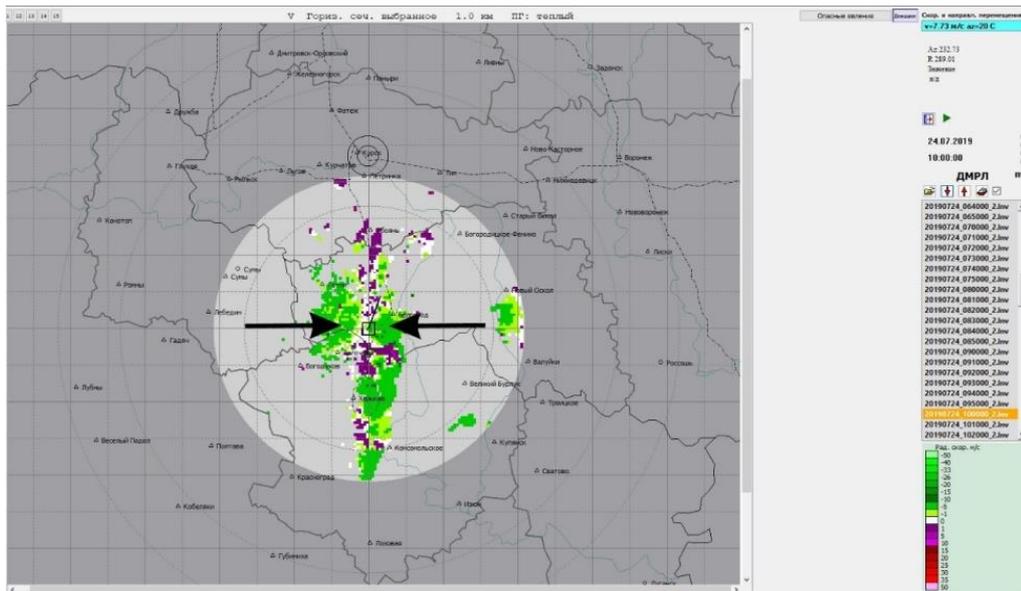


Рисунок 1.6. Карта радиальной доплеровской скорости на высоте 1 км.  
 ДМРЛ-С Белгород, 10:00 UTC 24 июля 2019г.

На данном примере видно столкновение воздушных масс прямо над ДМРЛ, где и произошло развитие конвективных явлений.

С помощью вторичных продуктов ДМРЛ появляется возможность рассмотреть зоны струйных течений на разных высотах, что может быть полезно для авиации, так как можно точно определить зоны течений. Для сравнения, была взята аэрологическая станция Курск (расстояние 144 км), так как на станции Харьков измерения в 12:00 UTC не проводятся.

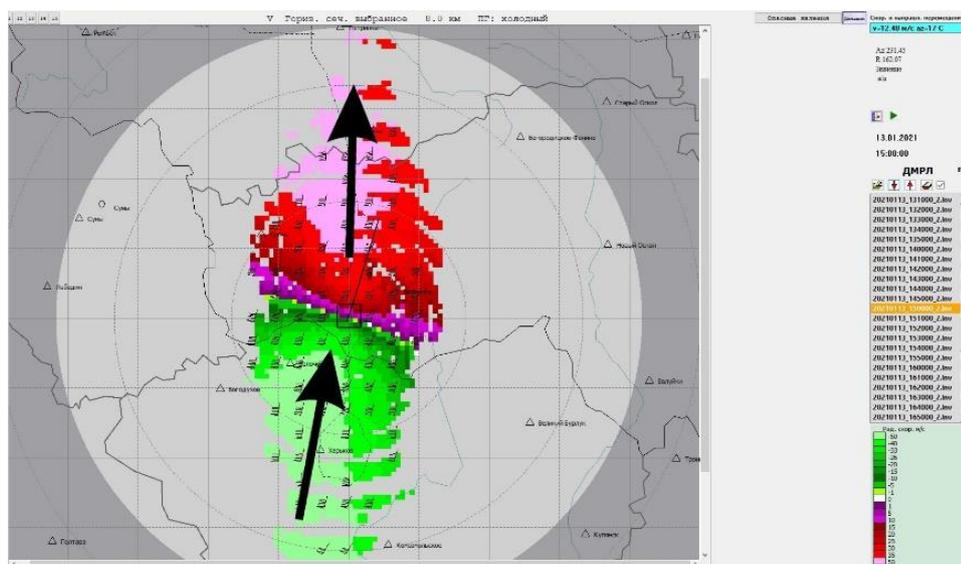


Рисунок 1.7. Карта радиальной доплеровской скорости на высоте 8 км.  
 ДМРЛ-С Белгород, 13 января 15:00 UTC.

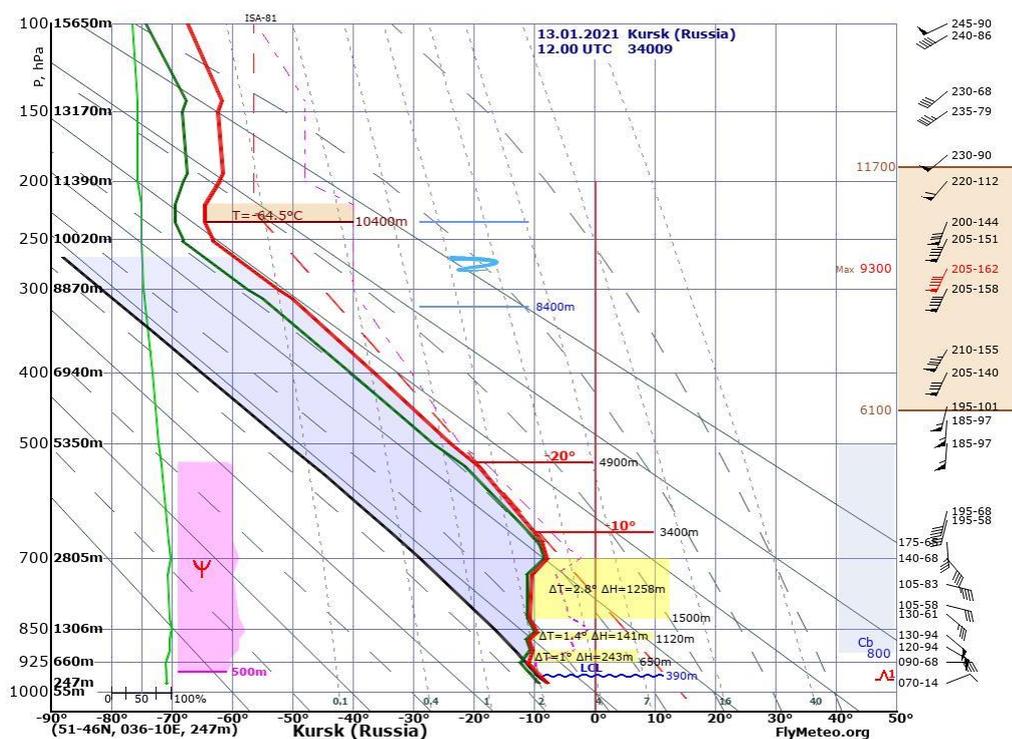


Рисунок 1.8. Аэрологическая диаграмма за 13 января 12:00 UTC, станция Курск.

Анализ радиальных скоростей по ДМРЛ преимущественно используется в авиационном прогнозе. Пилотам самолётов необходимо знать о сдвигах ветра при взлёте и посадке, а данные о струйных течениях помогут грамотно проложить маршрут и сэкономить, как время для пассажиров, так и уменьшить издержки для авиакомпании.

### 1.5. Радиолокационное измерение осадков.

Сумма накопленных осадков  $Q$  измеряется в мм слоя воды, выпавшей на единичную площадку за заданный интервал времени. Суммирование осадков на ДМРЛ-С предусмотрено за различные промежутки времени- 1, 3, 6, 12 и 24 часа, причём для каждого радиолокатора расчёт сумм осадков проводится в соответствии с их расположением по метеозонам. Так, расчёт 24- часовой (суточной) суммы осадков начинается с началом метеорологических суток и заканчивается по их окончании; расчёт первой 12-часовой суммы начинается с началом метеорологических суток, второй

12-часовой суммы - спустя 12 часов; суммы осадков за 6 и 3 ч привязываются к основным и промежуточным срокам.

Ливневые дожди с количеством выпавших осадков не менее 30 мм/ч, жидкие или смешанные осадки в количестве не менее 50 мм, выпавшие не более чем за 12 часов, а также снег не менее 20 мм, выпавший не более чем за 12 часов, относятся к опасным гидрометеорологическим явлениям.

Поэтому, анализируя карты интенсивности и суммы осадков за 1 час и за полусутки, можно оценить зоны с опасным количеством выпавших осадков.[1]

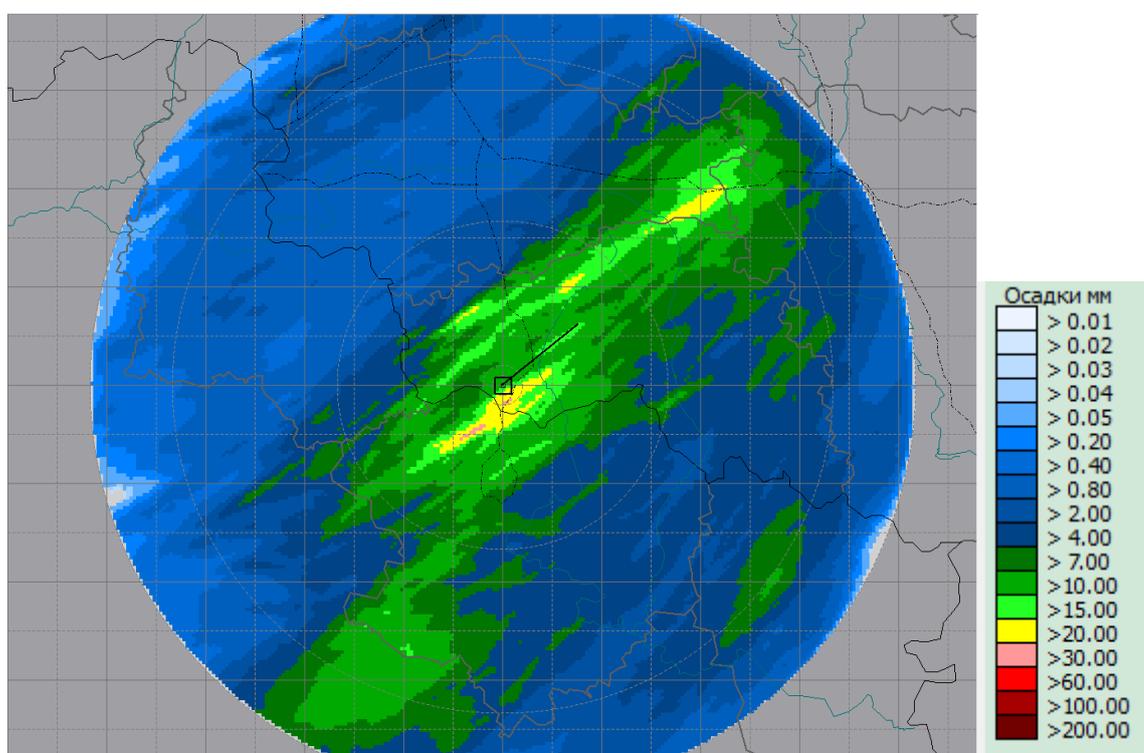


Рисунок 1.9. Карта сумм осадков за 12 часов. ДМРЛ-С Белгород, 18:00 UTC 8 июля 2019 года

Но также, как и в случае с другими изображениями доплеровского радара, при просмотре этого изображения необходимо проявлять некоторую осторожность, поскольку необходимо учитывать два основных фактора.

Во-первых, хотя радар отлично справляется с самокоррекцией, бывают случаи, когда радар не поверяется. Если радар сообщает о слишком сильном эхо, то оценки количества осадков будут завышенными, и наоборот, при слабом эхо радар недооценит количество осадков. Всегда проверяйте

ближайшие радары, чтобы убедиться, что они сообщают информацию, аналогичную той, которую просматривает ваш местный радар.

Во-вторых, град — отличный отражатель энергии. Грозы с градом завышают количество осадков, и чем крупнее градины, тем больше переоценка.

## ГЛАВА 2. Машинное обучение в задачах метеорологии

В метеорологии одна из основных задач – прогноз погоды. Для этого отлично подходят алгоритмы машинного обучения. В последние годы, с развитием алгоритмов глубокого обучения, становится возможным получать более точные модели для прогноза погоды. Главное отличие от гидродинамических моделей заключается в том, что для моделей глубокого обучения, то есть нейросетей, необходимы большие объёмы архивных данных. Данный подход базируется на пространственном и временном изменении метеорологических характеристик, выбор характеристик же подчиняется законам физики и знаниям о том, какие из них имеют влияние на целевую переменную (прогнозируемое событие).

К примеру, в технологии прогнозирования погоды – “Метеум”, которая используется в сервисе Яндекс.Погода для построения наиболее актуального прогноза для заданных координат.

Для выявления закономерностей и взаимосвязей между прогнозами моделей и реальной метеорологической обстановкой используется известный алгоритм машинного обучения Матрикснет, который принимает на вход специальным образом переработанные архивы прогнозов погоды и сопоставляет их с данными о реальной метеорологической обстановке. В качестве данных о реальной погоде используются наблюдения, полученные на тысячах профессиональных метеостанций по всему миру.

В результате таких сопоставлений получается формула корректировки прогноза, которая, в зависимости от метеорологической ситуации, составляет оптимальную комбинацию прогнозов из данных моделей. По своим собственным подсчетам, Яндекс выяснил, что им удалось достичь точности температурного расчета, на 35% превышающей результаты ближайшего конкурента. Однако на этом их исследования не остановились, осенью этого года Яндекс представил технологию гиперлокального краткосрочного прогноза осадков. Для повышения точности было принято решение –

добавить в математически сформированные прогнозы реальные данные с метеостанций и радиолокаторов. Метеорологический радиолокатор – самое последнее достижение науки и техники в области мониторинга. Таким образом, с точки зрения математики задача формулируется следующим образом: необходимо по последовательности радарных снимков и релевантной метеорологической информации предсказать будущие карты осадков с временным разрешением в 10 минут на 2 часа вперёд.

Технология гиперлокального краткосрочного прогноза осадков основана на сверточных нейронных сетях. Каждый сверточный слой сети представляет собой последовательный набор фильтров, автоматически выделяющих признаки из набора данных, подаваемых на вход. Совокупность таких слоёв позволяет выделить большое количество признаков, по которым можно судить о том, что будет происходить с осадками в выделенной области. Чтобы настроить фильтры сверточной сети и веса между ними, используется обучающая выборка достаточно большого объёма — сейчас она состоит более чем из 800000 кадров с каждого радара, прошедших предварительную обработку.

Данная технология позволяет узнать о распределении осадков во временном промежутке от двух часов назад до двух часов вперёд. Она получила название – технология наукастинга — краткосрочного гиперлокального прогноза осадков. [3]

## 2.1. Сверточная нейронная сеть.

Сверточная нейронная сеть (CNN), класс искусственных нейронных сетей, который стал доминирующим в различных задачах компьютерного зрения, вызывает интерес в различных областях, включая метеорологию. CNN предназначена для автоматического и адаптивного изучения пространственных иерархий объектов посредством обратного распространения с использованием нескольких слоёв, таких как слои свертки, объединяющие слои и полностью связанные слои. Знакомство с

концепциями и преимуществами, а также ограничениями CNN имеет важное значение для использования потенциала в метеорологии.

В последние годы возник огромный интерес к глубокому обучению [9]. Наиболее признанным алгоритмом среди различных моделей глубокого обучения является сверточная нейронная сеть (CNN), класс искусственных нейронных сетей, который является доминирующим методом в задачах компьютерного зрения. Знакомство с этой современной методологией поможет не только исследователям, которые применяют CNN для решения своих задач в радиологии и медицинской визуализации, но и клиническим радиологам, поскольку глубокое обучение может повлиять на их практику в ближайшем будущем. В этой статье основное внимание уделяется основным концепциям CNN и их применению к различным задачам радиологии, а также обсуждаются ее проблемы и будущие направления. Другие модели глубокого обучения, такие как рекуррентные нейронные сети для моделей последовательностей, выходят за рамки этой статьи.

Следующие термины будут использоваться. «Параметр» или «Вес» означает переменную, которая автоматически изучается в процессе обучения. «Гиперпараметр» относится к переменной, которую необходимо установить до начала процесса обучения. «Ядро» относится к наборам обучаемых параметров, применяемых в операциях свертки.

### 2.1.1. Общее объяснение сверточных нейронных сетей (CNN).

CNN — это математическая конструкция, которая обычно состоит из трех типов слоев (или строительных блоков): свертки, объединения и полносвязных слоев. Первые два слоя, свертки и объединения, выполняют извлечение признаков, тогда как третий, полносвязный слой, отображает извлеченные признаки в конечный результат, решает задачу, к примеру, классификации. Слой свертки играет ключевую роль в CNN, который состоит из набора математических операций, таких как свертка -

специализированный тип линейной операции. В цифровых изображениях значения пикселей хранятся в двумерной (2D) сетке, т. е. в массиве чисел (рис. 2.1), и в каждой позиции изображения применяется небольшая сетка параметров, называемая ядром, оптимизируемым экстрактором признаков. Это делает CNN очень эффективными для обработки изображений, поскольку функция может появиться в любом месте изображения. По мере того, как один слой передает свои выходные данные следующему слою, извлеченные функции могут иерархически и постепенно усложняться. Процесс оптимизации параметров, таких как ядра, называется обучением, которое выполняется, чтобы минимизировать разницу между выходными данными и метками с помощью алгоритма оптимизации, называемого обратным распространением и градиентным спуском.

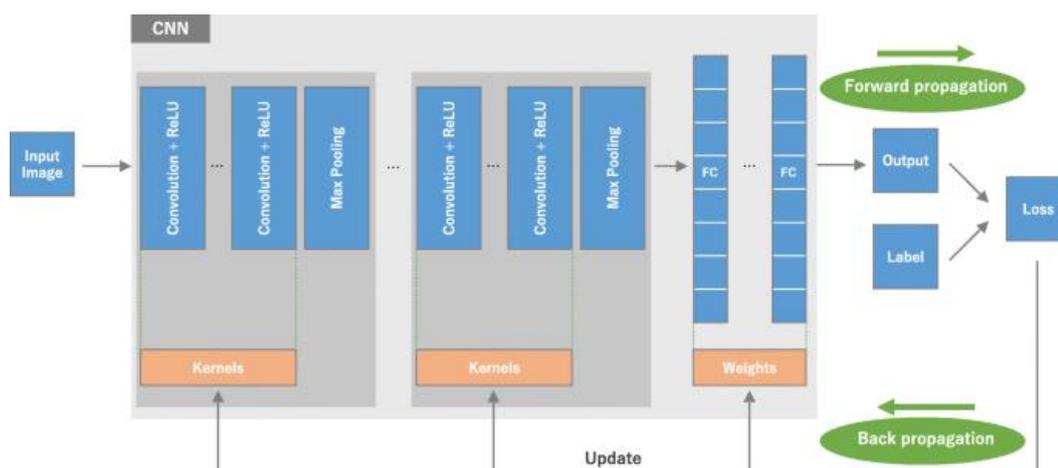


Рис. 2.1. Сверточная нейронная сеть, общий вид.

Обзор архитектуры сверточной нейронной сети (CNN) и процесса обучения. CNN состоит из набора нескольких строительных блоков: слоев свертки, слоев объединения (Max Pooling) и полностью связанных (FC) слоев. Производительность модели при определенных ядрах и весах рассчитывается с помощью функции потерь посредством прямого распространения в обучающем наборе данных, а обучаемые параметры, т. е. ядра и веса, обновляются в соответствии со значением потерь посредством



функции активации. Свертка — это специализированный тип линейной операции, используемый для извлечения признаков, когда небольшой массив чисел, называемый ядром, применяется ко входным данным, которые представляют собой массив чисел, называемый тензором. Поэлементное произведение между каждым элементом ядра и входным тензором вычисляется в каждой позиции тензора и суммируется для получения выходного значения в соответствующей позиции выходного тензора, называемого картой признаков (рис. 2.3а–с). Эта процедура повторяется с применением нескольких ядер для формирования произвольного количества карт признаков, которые представляют различные характеристики входных тензоров. Таким образом, разные ядра можно рассматривать как разные экстракторы признаков (рис. 2.3д). Двумя ключевыми гиперпараметрами, определяющими операцию свертки, являются размер и количество ядер. Первое обычно составляет  $3 \times 3$ , но иногда  $5 \times 5$  или  $7 \times 7$ . Последнее является произвольным и определяет глубину выходных карт объектов.

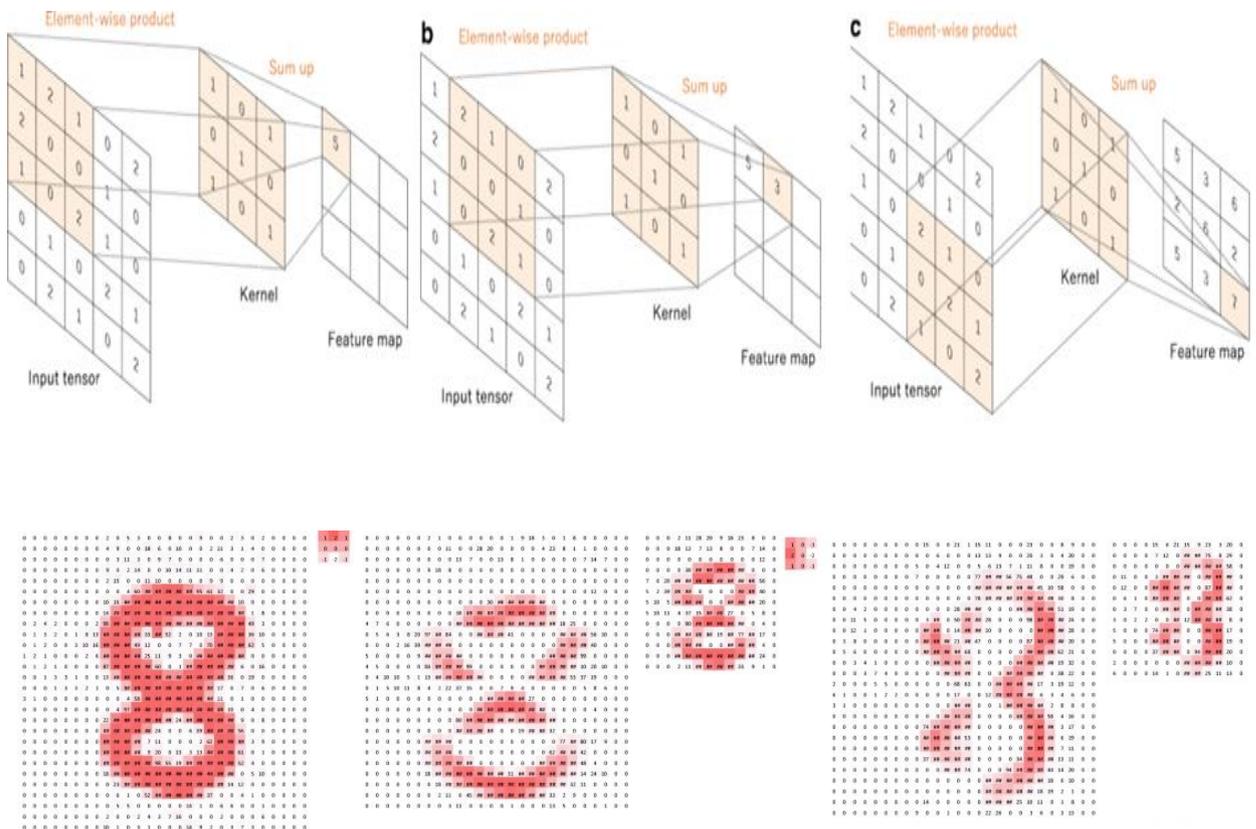


Рис 2.3. Визуализация свертки и пример работы на изображении

а–с Пример операции свертки с размером ядра  $3 \times 3$ , без заполнения и с шагом 1. Ядро применяется к входному тензору, а поэлементное произведение между каждым элементом ядра и входом тензор вычисляется в каждом месте и суммируется для получения выходного значения в соответствующем положении выходного тензора, называемого картой признаков. Показаны примеры того, как ядра в слоях свертки извлекают признаки из входного тензора. Несколько ядер работают как разные экстракторы признаков, такие как детектор горизонтальных краев (вверху), детектор вертикальных краев (посередине) и детектор контуров (внизу).

Описанная выше операция свертки не позволяет центру каждого ядра перекрывать самый внешний элемент входного тензора и уменьшает высоту и ширину выходной карты признаков по сравнению с входным тензором. Padding, как правило, zero padding (нулевое дополнение) — это метод решения этой проблемы, при котором строки и столбцы нулей добавляются с каждой стороны входного тензора, чтобы совместить центр ядра с самым внешним элементом и сохранить одно и то же измерение в плоскости через операция свертки (рис. 2.4). Современные архитектуры CNN обычно используют заполнение нулями, чтобы сохранить размеры в плоскости, чтобы применить больше слоев. Без заполнения нулями каждая последующая карта объектов будет уменьшаться после операции свертки.

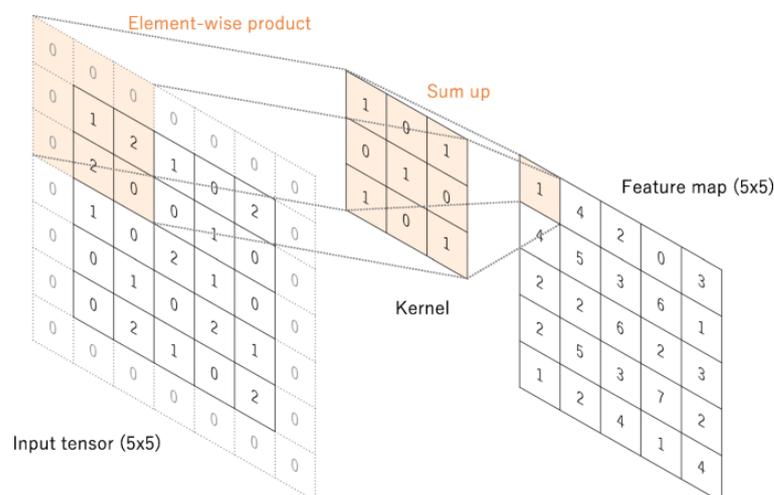


Рис. 2.4. Визуализация использования нулевого дополнения.

Ключевой особенностью операции свертки является распределение веса: ядра распределяются между всеми позициями изображения. Разделение веса создает следующие характеристики операций свертки: сохранение инвариантности локальных паттернов признаков, извлеченных путем трансляции ядер, поскольку ядра перемещаются по всем позициям изображения и обнаруживают изученные локальные паттерны, изучение пространственных иерархий паттернов признаков путем понижения дискретизации в сочетании с операцией объединения, что приводит к захвату все большего поля зрения и повышению эффективности модели за счет уменьшения количества параметров для изучения по сравнению с полностью подключенными нейронными сетями. Как описано ниже, процесс обучения модели CNN в отношении слоя свертки заключается в определении ядер, которые лучше всего подходят для данной задачи на основе данного набора обучающих данных. Ядра — это единственные параметры, которые автоматически изучаются в процессе обучения на сверточном слое; с другой стороны, размер ядер, количество ядер, заполнение и шаг — это гиперпараметры, которые необходимо установить до начала процесса обучения.

#### 2.1.2.2. Нелинейная функция активации.

Выходы линейной операции, такой как свертка, затем передаются через нелинейную функцию активации. Хотя гладкие нелинейные функции, такие как сигмоидная или гиперболическая тангенсная (тангенс) функция, использовались ранее, потому что они являются математическим представлением поведения биологического нейрона, наиболее распространенной нелинейной функцией активации, используемой в настоящее время, является выпрямленная линейная (ReLU), которая просто вычисляет функция:  $f(x) = \max(0, x)$  (рис. 2.5).

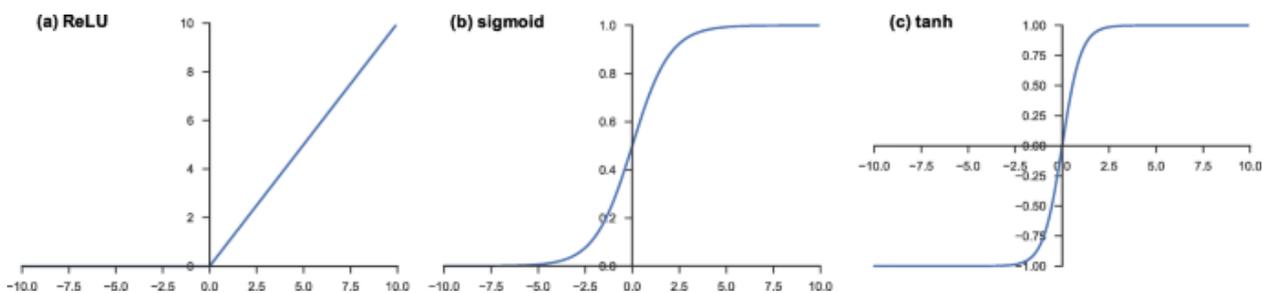


Рис. 2.5. Функции активации, обычно применяемые к нейронным сетям: а - выпрямленная линейная (ReLU), b – сигмоида, с - гиперболический тангенс (tanh).

### 2.1.2.3. Объединяющий слой.

Слой пула обеспечивает типичную операцию понижения дискретизации, которая уменьшает размерность карт объектов в плоскости, чтобы ввести трансляционную инвариантность к небольшим сдвигам и искажениям и уменьшить количество последующих обучаемых параметров. Следует отметить, что ни в одном из слоев пула нет обучаемого параметра, тогда как размер фильтра, шаг и заполнение являются гиперпараметрами в операциях пула, аналогичных операциям свертки.

Максимальное объединение - наиболее популярная форма операции объединения является максимальное объединение, которое извлекает исправления из входных карт объектов, выводит максимальное значение для каждого исправления и отбрасывает все остальные значения (рис. 2.6). На практике обычно используется максимальное объединение с фильтром размером  $2 \times 2$  с шагом 2. Это уменьшает плоскостное измерение карт объектов в 2 раза. В отличие от высоты и ширины, измерение глубины карт объектов остается неизменным.



каждой карте объектов, тогда как глубина карт объектов равна сохраняется. Эта операция обычно применяется только один раз перед полностью связанными слоями. Преимущества применения глобального среднего пула заключаются в следующем: уменьшает количество обучаемых параметров и позволяет CNN принимать входные данные переменного размера.

#### 2.1.2.4. Полностью связанный слой.

Карты выходных объектов конечного слоя свертки или объединения обычно сглаживаются, т. е. преобразуются в одномерный массив чисел (или вектор) и соединяются с одним или несколькими полносвязными слоями, также известными как плотные слои, в которых каждый вход связан с каждым выходом обучаемым весом. После того, как функции, извлеченные слоями свертки и прореженные слоями объединения, созданы, они сопоставляются подмножеством полносвязных слоев с окончательными выходными данными сети, такими как вероятности для каждого класса в задачах классификации. Последний полносвязный слой обычно имеет такое же количество выходных узлов, как и количество классов. За каждым полносвязным слоем следует нелинейная функция, такая как ReLU, как описано выше.

#### 2.1.2.5. Функция активации последнего слоя.

Функция активации, применяемая к последнему полносвязному слою, обычно отличается от остальных. Для каждой задачи необходимо выбрать соответствующую функцию активации. Функция активации, применяемая к задаче мультиклассовой классификации, представляет собой функцию softmax, которая нормализует выходные реальные значения из последнего полносвязного слоя в вероятности целевого класса, где каждое значение находится в диапазоне от 0 до 1, а сумма всех значений равна 1.

#### 2.1.3. Обучение нейронных сетей.

Обучение сети — это процесс поиска ядер в сверточных слоях и весов в полностью связанных слоях, которые минимизируют различия между выходными прогнозами и заданными метками истинности в наборе обучающих данных. Алгоритм обратного распространения — это метод, обычно используемый для обучения нейронных сетей, где важную роль играют функция потерь и алгоритм оптимизации градиентного спуска. Производительность модели при определенных ядрах и весах рассчитывается с помощью функции потерь посредством прямого распространения в наборе обучающих данных, а обучаемые параметры, а именно ядра и веса, обновляются в соответствии со значением потерь с помощью алгоритма оптимизации, называемого, среди прочего, обратным распространением и градиентным спуском. (Рисунок 2.1.).

#### 2.1.3.1. Функция потерь.

Функция потерь измеряет совместимость между выходными прогнозами сети посредством прямого распространения и заданными метками истинности. Обычно используемой функцией потерь для мультиклассовой классификации является перекрестная энтропия, тогда как среднеквадратическая ошибка обычно применяется для регрессии к непрерывным значениям. Тип функции потерь является одним из гиперпараметров и должен быть определен в соответствии с поставленными задачами.

#### 2.1.3.2. Градиентный спуск.

Градиентный спуск обычно используется в качестве алгоритма оптимизации, который итеративно обновляет обучаемые параметры, то есть ядра и веса сети, чтобы минимизировать потери. Градиент функции потерь указывает нам направление, в котором функция имеет наибольшую скорость роста, и каждый обучаемый параметр обновляется в отрицательном направлении градиента с произвольным размером шага, определяемым на

основе гиперпараметра, называемого скоростью обучения (рис. 17). Градиент математически представляет собой частную производную потерь по каждому обучаемому параметру, и одно обновление параметра формулируется следующим образом:

$$w := w - \alpha * \frac{\partial L}{\partial w},$$

где  $w$  обозначает каждый обучаемый параметр,  $\alpha$  обозначает скорость обучения, а  $L$  обозначает функцию потерь. Следует отметить, что на практике скорость обучения является одним из наиболее важных гиперпараметров, которые необходимо установить до начала обучения. На практике, по таким причинам, как ограничения памяти, градиенты функции потерь в отношении параметров вычисляются с использованием подмножества обучающего набора данных, называемого мини-пакетом, и применяются к обновлениям параметров. Этот метод называется стохастическим градиентным спуском (SGD), а размер мини-пакета также является гиперпараметром. Кроме того, было предложено и широко используется множество улучшений алгоритма градиентного спуска, таких как SGD с импульсом, RMSprop и Adam, хотя детали этих алгоритмов выходят за рамки этой статьи.

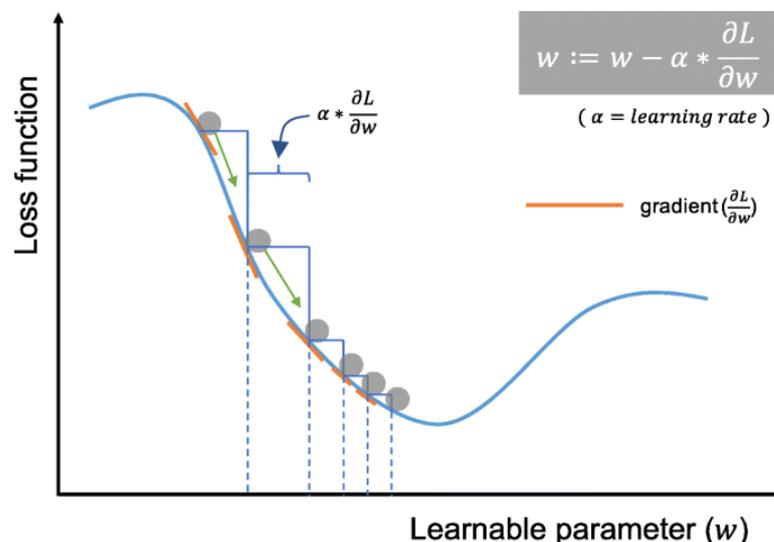


Рис. 2.7. Градиентный спуск.

Градиентный спуск — это алгоритм оптимизации, который итеративно обновляет обучаемые параметры, чтобы минимизировать потери, которые измеряют расстояние между выходным прогнозом и меткой истинности. Градиент функции потерь обеспечивает направление, в котором функция имеет наибольшую скорость роста, и все параметры обновляются в отрицательном направлении градиента с размером шага, определяемым на основе скорости обучения.

#### 2.1.4. Данные и разметка данных.

Данных и разметка данных являются наиболее важными компонентами исследований с применением глубокого обучения или других методов машинного обучения. Тщательный сбор данных и меток истинности, с помощью которых можно обучать и тестировать модель, является обязательным для успешного проекта глубокого обучения, но получение высококачественных маркированных данных может быть дорогостоящим и трудоемким.

Доступные данные обычно разбиваются на три набора: обучающий, проверочный и тестовый (рис. 2.8.), хотя есть и другие варианты, например перекрестная проверка. Учебный набор используется для обучения сети, где значения потерь рассчитываются посредством прямого распространения, а обучаемые параметры обновляются посредством обратного распространения. Набор проверки используется для оценки модели в процессе обучения, тонкой настройки гиперпараметров и выбора модели. Тестовый набор в идеале используется только один раз в самом конце проекта, чтобы оценить производительность окончательной модели, которая была настроена и выбрана в процессе обучения с наборами для обучения и проверки.

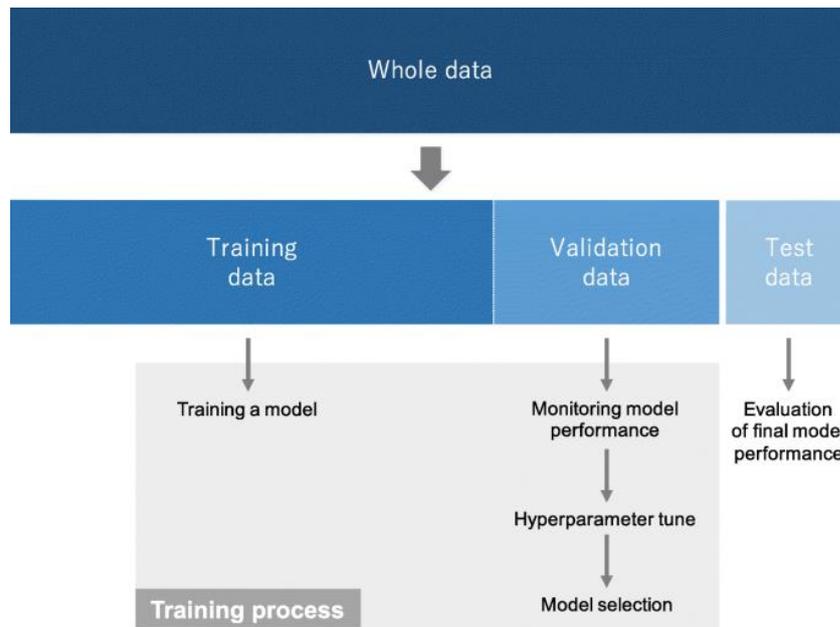


Рис. 2.8. Схема разделения данных для задач машинного обучения.

Необходимы отдельные наборы для проверки и тестирования, поскольку обучение модели всегда включает в себя точную настройку ее гиперпараметров и выполнение выбора модели. Поскольку этот процесс выполняется на основе производительности проверочного набора, некоторая информация об этом проверочном наборе просачивается в саму модель, т. е. Подгонка к проверочному набору, даже если модель никогда не обучалась на нем напрямую для обучаемых параметров. По этой причине гарантируется, что модель с точно настроенными гиперпараметрами в проверочном наборе будет хорошо работать в этом же проверочном наборе. Следовательно, для надлежащей оценки производительности модели необходим полностью невидимый набор данных, т. е. отдельный набор тестов, поскольку нас интересует производительность модели на никогда ранее не публиковавшихся данных, т. е. обобщаемость.

Стоит отметить, что термин «валидация» используется по-разному в области метеорологии и области машинного обучения. Как описано выше, в машинном обучении термин «валидация» обычно относится к этапу тонкой настройки и выбора моделей в процессе обучения. С другой стороны, в метеорологии «валидация» обычно означает процесс проверки

работоспособности прогностической модели, что аналогично термину «тест» в машинном обучении. Во избежание этой путаницы слово «набор для разработки» иногда используется вместо «набор для проверки».

#### 2.1.5. Переобучение.

Переобучение. Переобучение относится к ситуации, когда модель изучает статистические закономерности, характерные для обучающего набора, то есть в конечном итоге запоминает нерелевантный шум вместо изучения сигнала и, следовательно, хуже работает с последующим новым набором данных. Это одна из основных проблем в машинном обучении, поскольку переобученную модель нельзя использовать на ранее невиданные данные. В этом смысле набор тестов играет ключевую роль в правильной оценке производительности моделей машинного обучения, как обсуждалось в предыдущем разделе. Обычная проверка для распознавания переобучения заключается в отслеживании потерь и точности на обучающих и проверочных наборах (рис. 2.9.). Если модель хорошо работает на обучающем наборе по сравнению с проверочным набором, то модель, вероятно, была адаптирована к обучающим данным. Было предложено несколько методов минимизации переобучения: больше тренировочных данных, увеличение размерности данных, регуляризация (снижение веса, отсев), пакетная нормализация, уменьшение сложности архитектуры. Лучшее решение для уменьшения переобучения — получить больше обучающих данных. Модель, обученная на большем наборе данных, обычно лучше обобщает, что и достижимо в метеорологии. Другие решения включают в себя регуляризацию с отсевом или снижением веса, нормализацию пакетов и увеличение данных, а также снижение архитектурной сложности. Исключение (Dropout) — это недавно представленный метод регуляризации, при котором во время обучения случайно выбранным данным присваивается значение 0, так что модель становится менее чувствительной к конкретным весам в сети. Снижение веса, также называемое регуляризацией L2,

уменьшает переоснащение за счет штрафа за веса модели, так что веса принимают только небольшие значения. Пакетная нормализация — это тип дополнительного слоя, который адаптивно нормализует входные значения следующего слоя, снижая риск переобучения, а также улучшая градиентный поток через сеть, обеспечивая более высокую точность. скорости работы и снижения зависимости от инициализации. Увеличение размерности данных также эффективна для уменьшения переобучения, т. е. процесса модификации обучающих данных с помощью случайных преобразований, таких как переворачивание, перемещение, обрезка, вращение и случайное стирание, так что модель не будет видеть одни и те же входные данные вовремя обучающие итерации. Несмотря на эти усилия, по-прежнему существует проблема переобучения проверочного набора, а не обучающего набора, из-за утечки информации в процессе тонкой настройки гиперпараметров и выбора модели. Таким образом, отчет о производительности окончательной модели на отдельном (невидимом) тестовом наборе и, в идеале, на внешних наборах данных проверки, если применимо, имеет решающее значение для проверки обобщаемости модели.

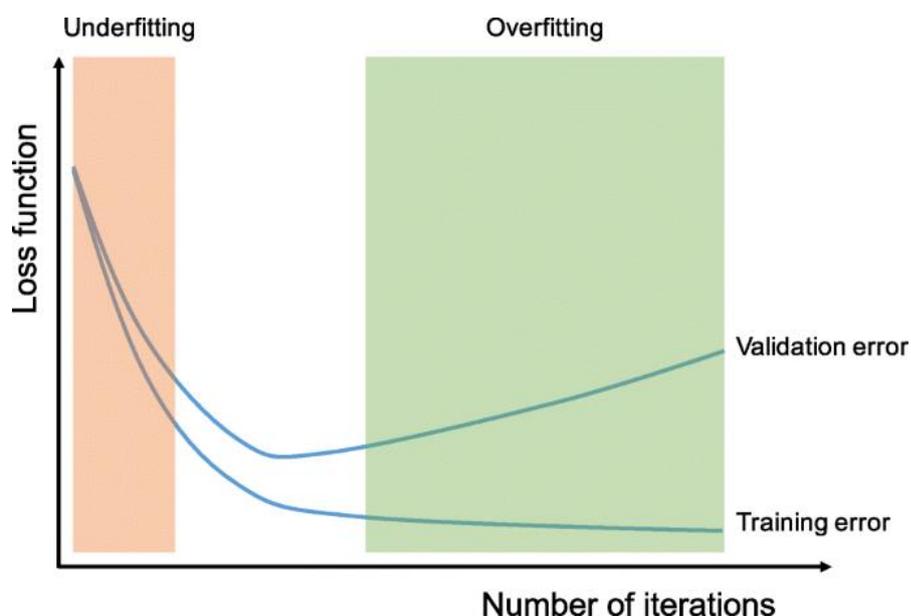


Рис. 2.9. График зависимости функции ошибки от количества итераций модели (Красная зона – недообучение, зелёная - переобучение)

Обычная проверка распознавания переобучения заключается в отслеживании потерь в обучающих и проверочных наборах во время итерации обучения. Если модель хорошо работает на обучающем наборе по сравнению с проверочным набором, то модель была адаптирована к обучающим данным. Если модель плохо работает как на обучающих, так и на проверочных наборах, то модель не соответствует данным. Хотя чем дольше сеть обучается, тем лучше она работает на тренировочном наборе, в какой-то момент сеть слишком хорошо подходит к обучающим данным и теряет способность к обобщению.

#### 2.1.6. Проблемы и дальнейшее развитие.

Хотя недавние достижения глубокого обучения были удивительными, все еще существуют проблемы с его применением в метеорологии.

Глубокое обучение считается черным ящиком, поскольку оно не оставляет контрольного следа для объяснения своих решений. Исследователи предложили несколько методов в ответ на эту проблему, которые дают представление о том, какие функции идентифицируются на картах функций, называемых визуализацией функций, и какая часть входных данных отвечает за соответствующий прогноз, называемый атрибуцией. Для визуализации признаков Zeiler и Fergus описали способ визуализации карт признаков, где первые слои идентифицируют небольшие локальные паттерны, такие как края или круги, а последующие слои постепенно объединяют их в более значимые структуры. Для атрибуции Zhou et al. предложил способ создания грубых карт локализации, называемых картами активации классов (CAM), которые локализуют важные области во входных данных, используемых для предсказания (рис. 2.10.). С другой стороны, стоит отметить, что исследователи недавно заметили, что глубокие нейронные сети уязвимы для крупных примеров, которые представляют собой тщательно подобранные входные данные, заставляющие сеть изменять выходные данные без видимого для человека изменения (рис. 2.11.). Хотя влияние грубых

примеров в метеорологической сфере неизвестно, эти исследования показывают, что то, как искусственные сети видят и предсказывают, отличается от того, как это делаем мы. Исследование уязвимости глубоких нейронных сетей в метеорологии имеет решающее значение, поскольку применение глубокого обучения требует чрезвычайной надежности для возможного использования в авиации по сравнению с относительно тривиальными методическими прогнозами.



Рис. 2.10. Пример карты активации класса (CAM)

Пример карты активации класса (CAM). Сеть CNN, обученная на ImageNet, классифицировала левое изображение как «пирс моста». Тепловая карта для категории «мостовой пирс», сгенерированная методом Grad-CAM, наложена (правое изображение), что указывает на области дискриминационного изображения, используемые CNN для классификации.



Рис. 2.11. Пример влияния шума при классификации изображения.

Состязательный пример, продемонстрированный Гудфеллоу. Сеть классифицировала объект на левом изображении как «панду» с достоверностью 57,7%. Добавив очень небольшое количество тщательно сконструированного шума (среднее изображение), сеть ошибочно классифицировала объект как «гиббон» с достоверностью 99,3% на правильном изображении без видимых изменений для человека.

Хотя существует несколько методов, которые облегчают обучение на небольших наборах данных, как описано выше, по-прежнему необходимы хорошо обработанные метеорологические наборы данных, поскольку большинство заметных достижений глубокого обучения обычно основаны на очень больших объемах данных. Целью больших наборов метеорологических данных является потенциал для повышения обобщаемости и минимизации переобучения, как обсуждалось ранее. Кроме того, когда такие наборы данных станут доступны, вероятно, можно будет предложить специализированные предварительно обученные метеорологические нейронные сети, которые могут способствовать исследованиям в области глубокого обучения метеорологии.

#### 2.1.7. Вывод о сверточных нейронных сетях.

Сверточные нейронные сети (CNN) добились удивительных успехов в различных областях, включая метеорологические исследования, и возрос интерес к наукастингу как для авиации, так и для общего использования. Хотя глубокое обучение стало доминирующим методом в различных сложных задачах, таких как классификация изображений и обнаружение объектов, оно не является панацеей. Знание ключевых концепций и преимуществ CNN, а также ограничений глубокого обучения необходимо для того, чтобы использовать его в метеорологических исследованиях с целью улучшения краткосрочных прогнозов.

## 2.2. U-NET сверточная нейронная сеть.

Существует широкое согласие с тем, что для успешного обучения глубоких сетей требуется большой объём данных для обучения. U-NET представляет собой сеть и стратегию обучения, основанную на активном использовании данных для более эффективного использования последних. Архитектура состоит из сужающегося пути для захвата контекста и симметричного расширяющегося пути, обеспечивающего точную локализацию. Такая сеть может быть обучена от начала до конца на небольшом количестве изображений и превосходит сверточную нейронную сеть. Так же сеть быстрее обучается, чем классическая CNN. Сегментация изображения 512x512 занимает меньше секунды на графическом процессоре последнего поколения.

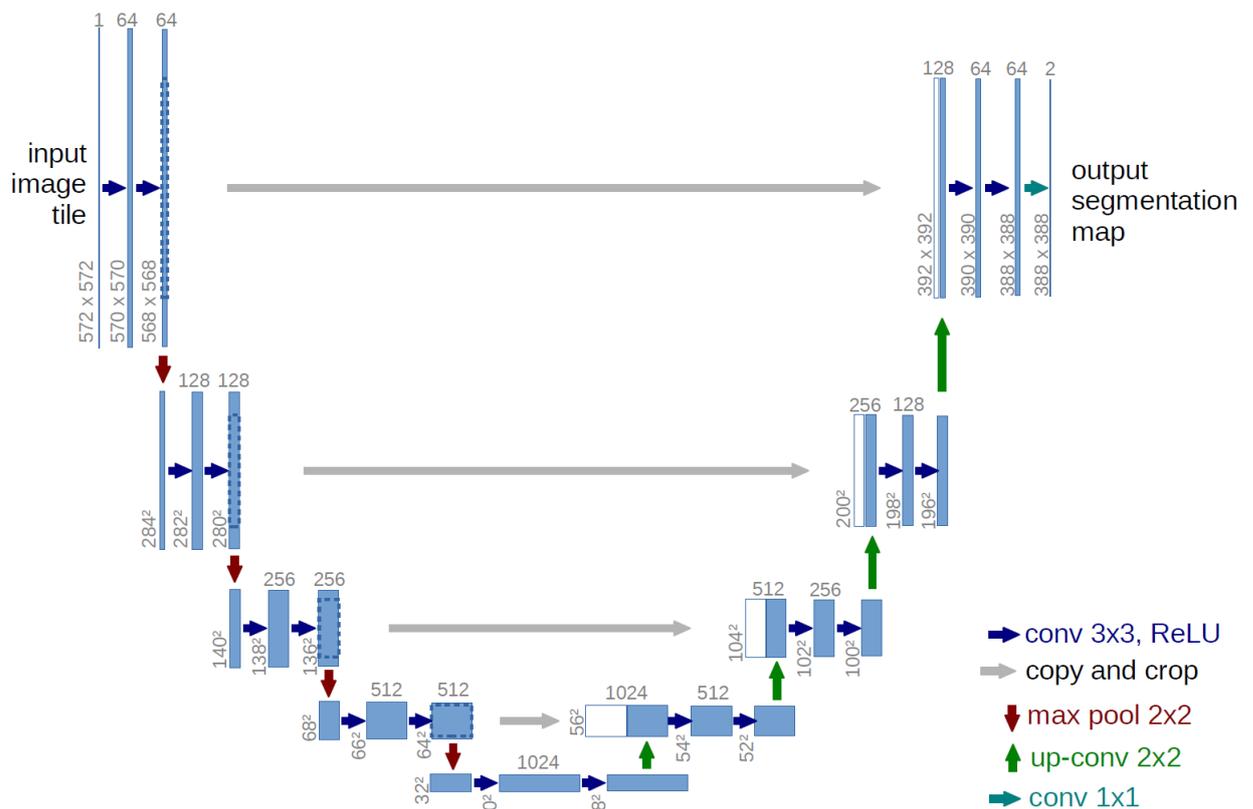


Рис. 2.12. Архитектура U-net.

Каждое синее поле соответствует многоканальной карте объектов.

Количество каналов указано в верхней части. Размер x-y указан в нижнем

левом углу поля. Белые прямоугольники представляют собой скопированные карты объектов. Стрелки обозначают различные операции.

Классическая сверточная нейронная сеть недостаточно хороша в данной задаче, поэтому стоит обратить внимание на более элегантную архитектуру, так называемой «полная сверточная нейронная сеть». Модификация и расширение этой архитектуры, чтобы она работала с очень небольшим количеством обучающих изображений и давала более точную сегментацию. Основная идея состоит в том, чтобы дополнить обычную сокращающуюся сеть последовательными слоями, где операторы объединения заменяются операторами повышающей дискретизации. Следовательно, эти слои увеличивают разрешение вывода. Чтобы локализовать признаки высокого разрешения из сокращающегося пути сочетаются с восстановленными выходными изображениями. Последующий слой свертки может затем научиться собирать более точные выходные данные на основе этой информации.

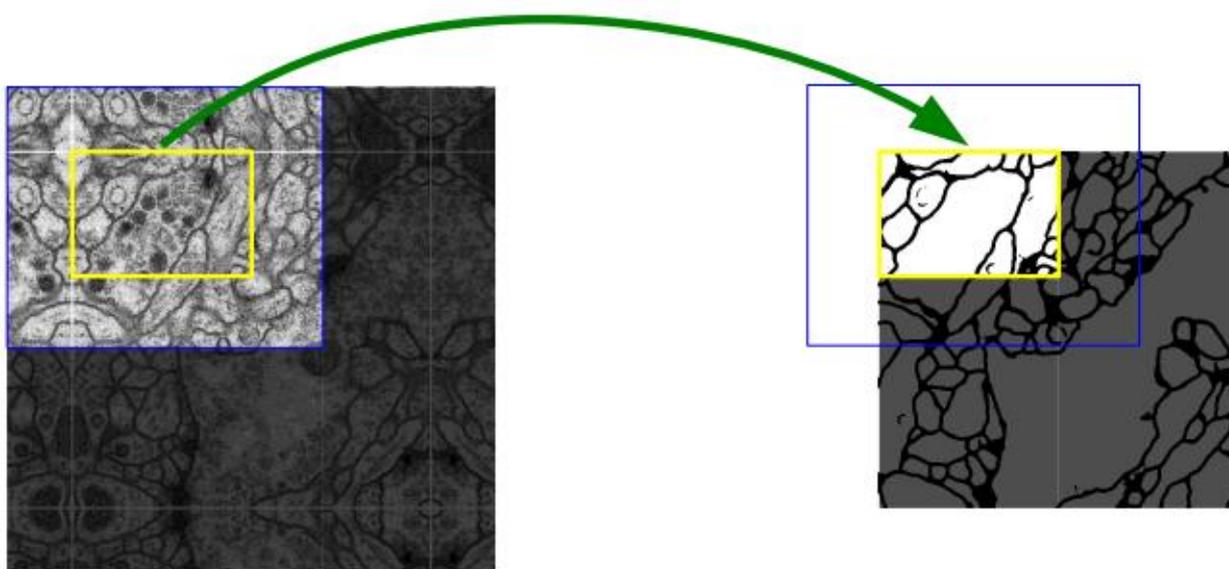


Рис. 2.13. Стратегия перекрывающихся плиток.

Стратегия перекрывающихся плиток для бесшовной сегментации произвольных больших изображений. Для прогнозирования сегментации в желтой области в качестве входных данных требуются данные изображения в

синей области. Отсутствующие входные данные экстраполируются путем зеркального отображения

Одно важное изменение в этой архитектуре заключается в том, что в части повышающей дискретизации также есть большое количество функциональных каналов, которые позволяют сети распространять контекстную информацию на уровне с более высоким разрешением. Как следствие, путь расширения более или менее симметричен пути сокращения и дает U-образную архитектуру. Сети не имеют полностью связанных слоев и используют только действительную часть каждой свертки, т. е. карта сегментации содержит только те пиксели, для которых во входном изображении доступен полный контекст. Эта стратегия позволяет плавно сегментировать произвольно большие изображения с помощью стратегии перекрывающихся фрагментов. Чтобы предсказать пиксели в граничной области изображения, отсутствующий контекст экстраполируется путем зеркального отражения входного изображения. Эта стратегия листов важна для применения сети к большим изображениям, поскольку в противном случае разрешение будет ограничено памятью графического процессора.

Использование избыточной размерности данных, применяя упругие деформации к имеющимся обучающим изображениям позволяет сети научиться инвариантности к таким деформациям без необходимости видеть эти преобразования в аннотированном корпусе изображений. Это особенно важно в биомедицинской сегментации, поскольку деформация раньше была наиболее распространенным изменением ткани, и реалистичные деформации можно эффективно моделировать.

### 2.2.1. Архитектура сети.

Архитектура сети показана на рисунке 2.12. Она состоит из сужающегося пути (слева) и расширяющегося пути (справа). Контрактный путь соответствует типичной архитектуре сверточной сети. Он состоит из

повторного применения двух сверток  $3 \times 3$  (сверток без дополнений), за каждой из которых следует выпрямленная линейная единица (ReLU) и операция максимального объединения  $2 \times 2$  с шагом 2 для понижения дискретизации. На каждом шаге понижения дискретизации мы удваиваем количество функциональных каналов. Каждый шаг расширенного пути состоит из повышающей дискретизации карты объектов, за которой следует свертка  $2 \times 2$  (up-convolution), которая вдвое уменьшает количество каналов признаков, конкатенация с соответствующим образом обрезанной картой признаков из сокращающегося пути и два  $3 \times 3$  свертки, за каждой следует ReLU. Обрезка необходима из-за потери граничных пикселей в каждой свертке. На последнем слое свертка  $1 \times 1$  используется для сопоставления каждого 64-компонентного вектора признаков с желаемым количеством классов. В Всего сеть имеет 23 сверточных слоя.

Чтобы обеспечить плавное разбиение выходной карты сегментации на листы (см. рис. 2.13.), важно выбрать размер входного фрагмента таким образом, чтобы все операции максимального объединения  $2 \times 2$  применялись к слою с четным размером  $x$  и  $y$ .

### 2.2.2. Обучение сети.

Входные изображения и соответствующие им карты сегментации используются для обучения сети с реализацией Sae со стохастическим градиентным спуском. Из-за незаполненных сверток выходное изображение меньше входного на постоянную ширину границы. Чтобы свести к минимуму накладные расходы и максимально использовать память графического процессора, мы предпочитаем большие входные фрагменты большому размеру пакета и, следовательно, уменьшаем пакет до одного изображения. Соответственно, мы используем высокий импульс (0,99), чтобы большое количество ранее просмотренных обучающих выборок определяло обновление на текущем шаге оптимизации.

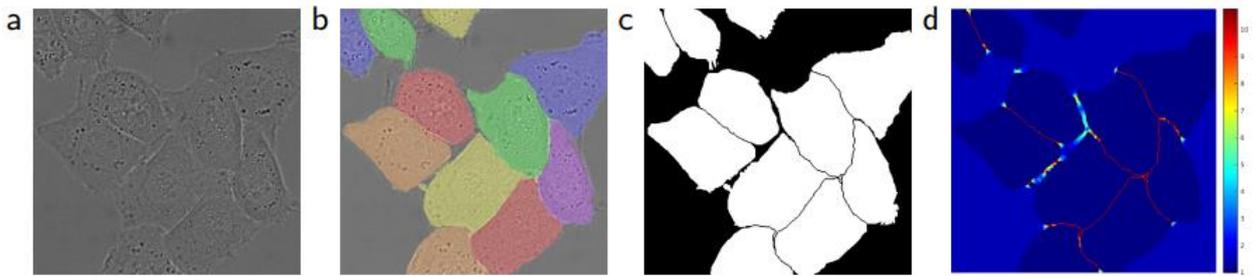


Рис. 2.14. Клетки HeLa на стекле, полученные с помощью ДИК-микроскопии (дифференциально-интерференционного контраста).

На выше представленном рисунке (a) необработанное изображение. (b) наложение с наземной сегментацией правды. Разные цвета указывают на разные экземпляры клеток HeLa. (c) сгенерированная маска сегментации (белая: передний план, черная: фон). (d) карта с попиксельной потерей веса, чтобы заставить сеть изучить граничные пиксели.

Мы предварительно вычисляем карту весов для каждой сегментации, чтобы компенсировать различную частоту пикселей из определенного класса в наборе обучающих данных и заставить сеть изучить небольшие границы разделения, которые мы вводим между соприкасающимися ячейками

В глубоких сетях с множеством сверточных слоев и различными путями в сети чрезвычайно важна хорошая инициализация весов. В противном случае одни части сети могут давать чрезмерную активацию, в то время как другие части никогда не вносят вклад. В идеале начальные веса должны быть адаптированы таким образом, чтобы каждая карта признаков в сети имела приблизительно единичную дисперсию. Для сети с нашей архитектурой (чередование слоев свертки и ReLU) это может быть достигнуто путем получения начальных весов из распределения Гаусса со стандартным отклонением  $\frac{1}{\sqrt{N}}$ , где  $N$  обозначает количество входящих узлов одного нейрона. Например, для свертки  $3 \times 3$  и 64 функциональных каналов в предыдущем слое  $N = 9 \cdot 64 = 576$ .

### 2.2.3. Увеличение размерности данных.

Увеличение размерности данных необходимо для обучения сети желаемым свойствам инвариантности и надежности, когда доступно только несколько обучающих выборок. В случае для микроскопических изображений нам в первую очередь нужна инвариантность к сдвигу и вращению, а также устойчивость к деформациям и вариациям уровня серого. Особенно случайные упругие деформации обучающих выборок кажутся ключевой концепцией для обучения сети сегментации с очень небольшим количеством аннотированных изображений. Мы генерируем гладкие деформации, используя случайные векторы смещения на грубой сетке 3 на 3. Смещения взяты из распределения Гаусса со стандартным отклонением 10 пикселей. Затем вычисляются попиксельные смещения с использованием бикубической интерполяции. Выпадающие слои в конце пути сокращения выполняют дальнейшее неявное увеличение данных.

#### 2.2.4. Вывод.

Архитектура U-NET обеспечивает очень хорошую производительность в самых разных приложениях. Благодаря аугментации данных эластичной деформаций, требуется очень мало изображений, а время обучения достаточно для использования на домашнем графическом процессоре. Мы предоставляем полную реализацию на основе Caffe и обученные сети. Архитектуру U-NET можно легко применить для многих других задач, таких как прогнозирование перемещения облаков с помощью данных доплеровского метеорологического локатора.

### ГЛАВА 3. Построение и использование U-NET сверточной нейронной сети для краткосрочного прогнозирования передвижения облачности по данным доплеровского метеорологического локатора.

В задачах машинного обучения есть определённый план работы с данными. Он включает в себя 4 пункта. Первым шагом является постановка задачи, затем следует обзор и обработка данных для модели. Третий шаг – выбор подходящей модели, её настройка и тестирование. В заключении, визуализируются полученные данные и дальнейшее использование модели в прикладных задачах.

#### 3.1. Постановка задачи.

Грамотная постановка задачи необходима для дальнейшей работы в рамках машинного обучения. Задачей является прогнозирование перемещения облачности по данным ДМРЛ. Такие проекты уже были реализованы на известных сайтах прогноза погоды, к примеру Яндекс.Погода. Следует рассмотреть данный пример на рисунке 3.1.

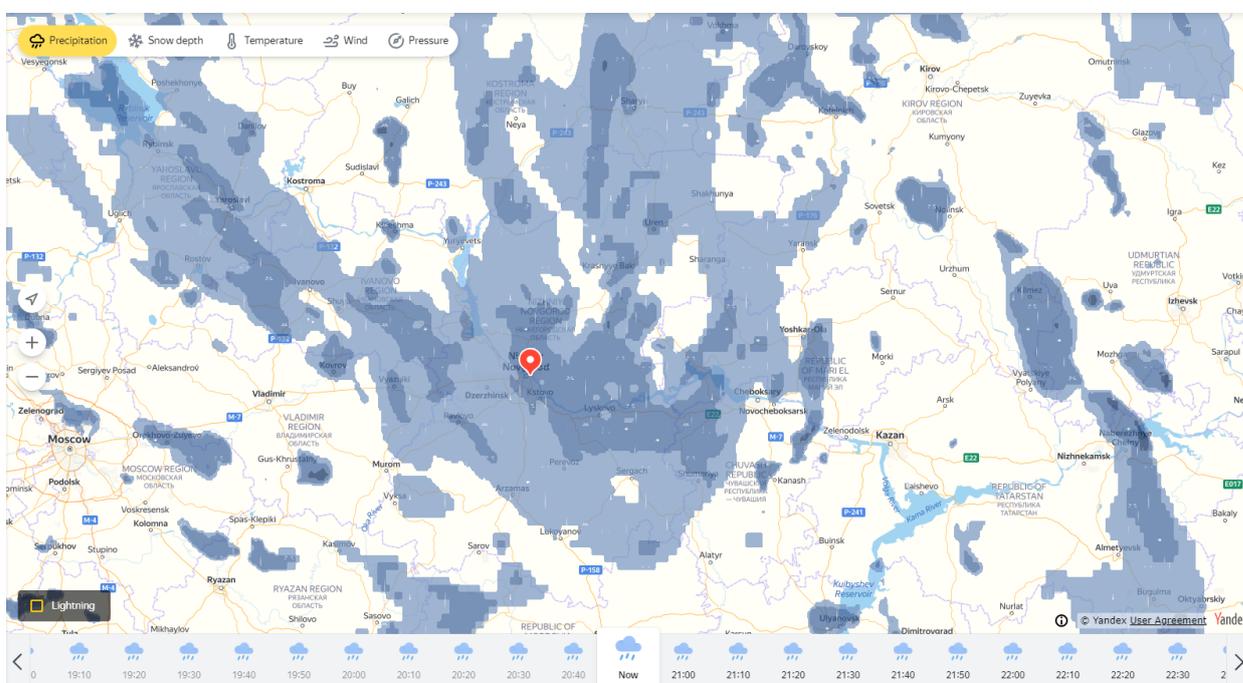


Рисунок 3.1. Изображение облачности над территорией Европейской части России в 20:50 по московскому времени 27 мая 2022 г.

На данном изображении отмечены зоны с осадками и их интенсивностью. Также имеется возможность просмотреть изображения в следующие 2 часа с дискретностью 10 минут с прогнозом перемещения зон осадков.

Целью данной работы является создать похожие изображения и спрогнозировать перемещение конвективной облачности во времени с помощью U-NET сверточной нейронной сети.

### 3.2. Обзор и обработка данных.

В машинном обучении предобработка данных является важным этапом. Данные нужно преобразовать в формат, который будет более удобным для дальнейшей работы с ними как пользователю, так и алгоритмам.

Для обработки данных используется код на языке программирования Python, который удобен из-за простоты синтаксиса и определённых библиотек, специально созданных для удобной работы с данными и реализации алгоритмов машинного обучения.

Для обработки данных необходимы библиотеки: NumPy и pandas. Быстрые и универсальные концепции векторизации, индексации NumPy сегодня фактически являются стандартами массивных вычислений. Так же данная библиотека предлагает комплексные математические функции, генераторы случайных чисел, процедуры линейной алгебры, преобразования Фурье и многое другое. Pandas — это библиотека Python для обработки и анализа структурированных данных, её название происходит от «panel data» («панельные данные»). Панельными данными называют информацию, полученную в результате исследований и структурированную в виде таблиц. Для работы с такими массивами данных и создан Pandas.

Данные предоставлены ООО “Институт радарной метеорологии”, ИРАМ. В распоряжении имеется данные о конвективной облачности за 2015 год ДМРЛ Пулково. Содержание файлов следующее (таблица 3.1.):

Таблица 3.1.

Данные ДМРЛ Пулково за 2015 по конвективным осадкам.

	cell_id	dt_stamp	source_id	ebottom	etop	phen	zh	vradh	wradh	zdr	..
0	776286036	2021-05-08 01:40:00	26065	NaN	7750.0	24	{29.0,29.0,29.0,29.0,14.0,14.0,14.0,14.0,6.0,6...	NaN	NaN	NaN	..
1	780389628	2021-05-12 03:20:00	26065	NaN	4500.0	24	{28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,...	NaN	NaN	NaN	..
2	780389629	2021-05-12 03:20:00	26065	NaN	4500.0	24	{28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,...	NaN	NaN	NaN	..
3	780389630	2021-05-12 03:20:00	26065	NaN	4500.0	24	{28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,...	NaN	NaN	NaN	..
4	780389631	2021-05-12 03:20:00	26065	NaN	4500.0	24	{28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,28.0,...	NaN	NaN	NaN	..

На данный момент в данных много пустых значений. Первая причина этому – ДМРЛ Пулково не имеет двойной поляризации из-за чего часть первичной радиолокационной информации отсутствует. Вторая причина – количество уровней, на котором был получен сигнал от облака редко достигает максимума и значений на большинстве уровней нет. Из-за такой проблемы, следует более простое решение. Использование данных только о верхней границе облачности будет достаточным для прогноза перемещения облачности.

В полученных данных, первый столбец — это номер ячейки, далее время, код станции, нижняя граница облачности (данные отсутствуют), верхняя граница облачности, закодированное явление погоды и значение отражаемости по слоям.

Для создания изображений верхней границы облачности, необходимо создать матрицу с номерами ячеек привязанную к координатам. Для раскодирования, используется предоставленный ИРАМ вариант.

Каждая ячейка имеет уникальный идентификационный номер, который рассчитывается по следующей формуле:

$$\text{Cell\_ID} = 1 + \text{int} [(\text{lon\_sec} + 180*60*60)/\text{lon\_step\_sec}] + \{\text{int}[(\text{lat\_sec} + 90*60*60)/\text{lat\_step\_sec}]\}*(360*60*60/\text{lon\_step\_sec}),$$

где  $\text{lon\_sec}$  - долгота базовой точки ячейки в секундах,  $\text{lat\_sec}$  - широта базовой точки ячейки в секундах,  $\text{lon\_step\_sec}$  - шаг сетки по долготу в секундах (=30),  $\text{lat\_step\_sec}$  - шаг сетки по широте в секундах (=30).

Для обработки данных было выбрано 5 миллионов строк из полученных данных, хотя в данных в общей сумме 400 миллионов строк, возможным представилось обработать только немногим более одного процента. Это было вызвано из-за нехватки мощностей домашнего персонального компьютера для преобразования данных в изображения.

В полученных данных, было выявлено 1427 моментов во времени. Установлен размер матрицы исходя из покрытия радара – 481 на 961 пиксель. Для удобства работы, матрица была сокращена до 480 на 960. Объединяя полученные данные, получили матрицу с размерностью – 1427, 480, 960. Далее полученную матрицу необходимо заполнить значениями о верхней границе облачности. Для успешной работы модели, был выбран итоговый размер изображения 256, 256 пикселей, так как модель рассчитана на квадратные изображения.

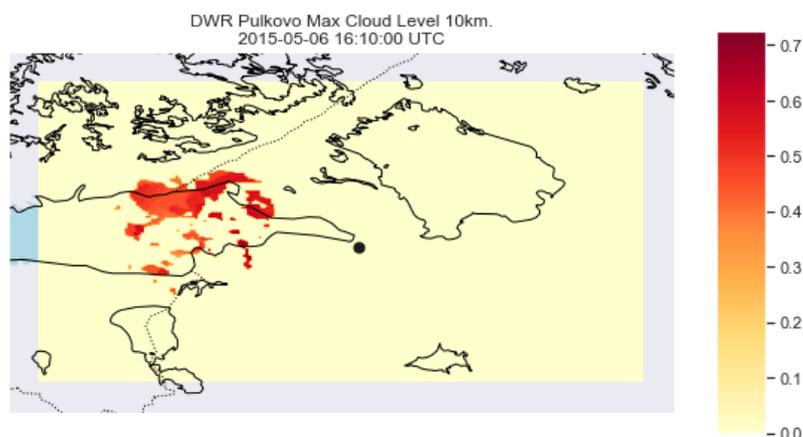


Рисунок 3.2. Полученное изображение после обработки данных.

Для обучения модели все данные были разделены по 9 изображений, пять из которых предназначаются для обучения, а 4 как метки, которые будут прогнозироваться. В прикладной задаче допустим, что 5 изображений – это прошедшее фактическое состояние облачности, а оставшиеся 4 – положение в будущем.

### 3.3. Выбор модели, её настройка и тестирование.

О модели U-NET говорилось во 2 главе. Выбор данной модели обусловлен её универсальностью для работы с изображениями и небольшим объёмом данных для обучения.

Сверточная нейронная сеть U-NET была реализована с помощью библиотеки Tensorflow (Google), в частности с использованием пакета Keras. В качестве оптимизатора был выбран стохастический градиентный спуск, функция потерь – бинарная кросс энтропия, метрика – бинарная точность.

Кросс-энтропия (или логарифмическая функция потерь –  $\log \text{loss}$ ): Кросс-энтропия измеряет расхождение между двумя вероятностными распределениями. Если кросс-энтропия велика, это означает, что разница между двумя распределениями велика, а если кросс-энтропия мала, то распределения похожи друг на друга. Бинарная точность измеряет как часто прогноз совпадает с фактическим положением.

Модель была обучена на данных для обучения 20 раз (эпох, epochs). После чего были получены следующие значения: функция потерь составила – 0.0151, оценка бинарной точности составила 0.9835. Такие завышенные оценки можно оправдать тем, что имеется небольшой объём данных за один месяц и модель точно определяет перемещение в рамках полученных данных. Далее приведены графики изменения функции ошибки и метрики по эпохам.

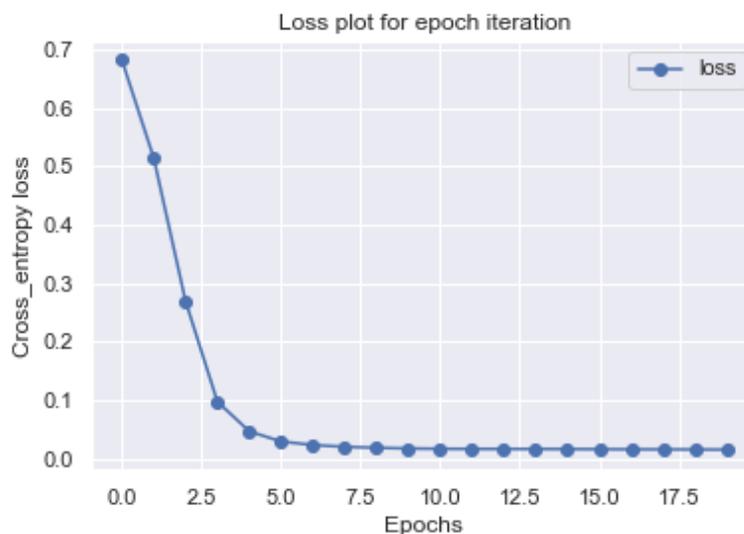


Рисунок 3.3. График изменения функции ошибки по эпохам.

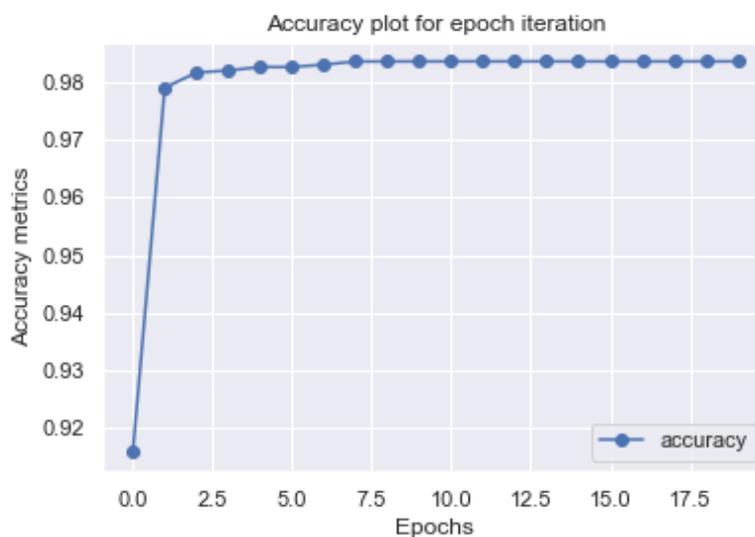


Рисунок 3.4. График изменения метрики по эпохам.

По графикам очевидно, что после 5 эпохи, значения ошибки перестали значительно уменьшаться, а на графике точности, после 3 итерации. Но полученные оценки нужно протестировать на данных для проверки.

Данные для проверки – миллион строк данных, 659 моментов во времени мая 2015 года. Для подготовки этих данных использовался тот же алгоритм, что и для данных для обучения. Были полученные следующие оценки: функция потерь составила – 0.0061, оценка бинарной точности

составила 0.993. Данные оценки говорят о большей успешности модели, чем на обучающей выборке, что является отличным результатом.

### 3.4. Визуализация результатов.

Для визуальной оценки работы модели, необходимо вывести изображения из выборки для тестирования и полученного прогноза, рисунки 3.5. и 3.6. соответственно.

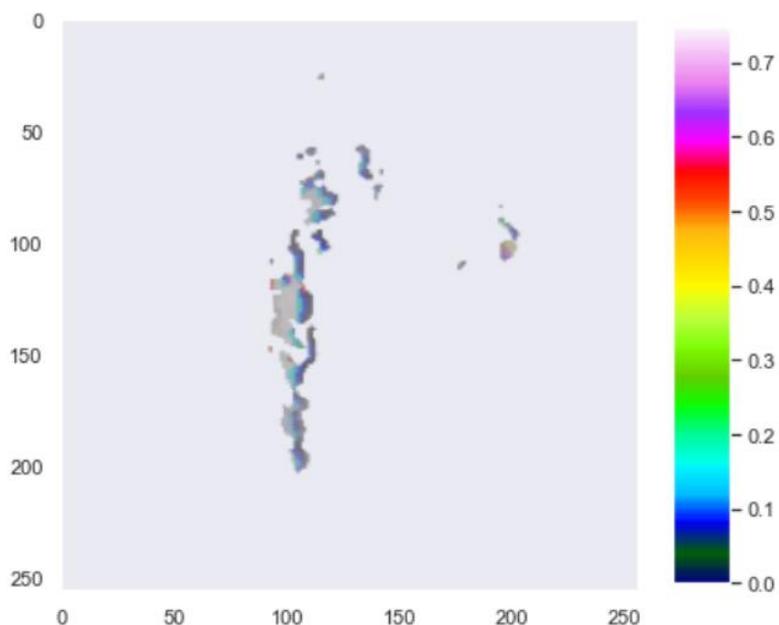


Рисунок 3.5. Изображение из тестовой выборки верхней границы облачности, 10 км.

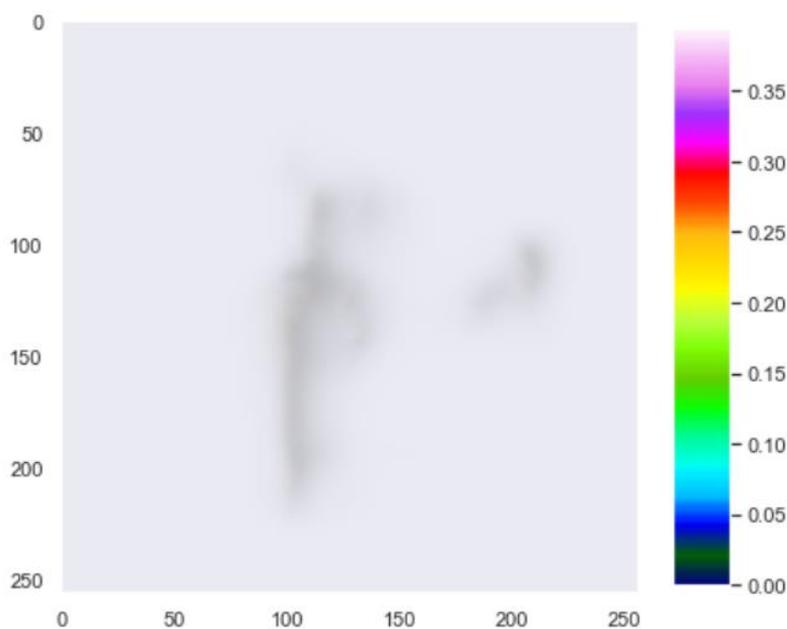


Рисунок 3.6. Изображение прогноза верхней границы облачности, 10 км.

На изображении прогноза видны границы облачности, и они совпадают с фактическим, но значения высоты занижены потому, что для обучения модели были выбраны бинарные гиперпараметры, которые учитывают только расположение облачности, но не уровень верхней границы.

## ЗАКЛЮЧЕНИЕ

Данные доплеровского метеорологического лоатора являются крайне полезным ресурсом в нынешнее время для краткосрочного прогнозирования атмосферных явлений. Машинное обучение – новый подход к прогнозированию, который не распространён в метеорологии, но уже очевидно, что потенциал использования безграничен. В совокупности с гидродинамическими моделями алгоритмы машинного обучения выводят прогнозирование атмосферных явлений на новый уровень.

В этой работе были рассмотрены такие алгоритмы как сверточные нейронные сети и их применение в метеорологии. Модель U-NET крайне удобный инструмент для работы с изображениями, полученными с ДМРЛ и и дальнейшее широкое применение этих алгоритмов лишь вопрос времени.

Модель, представленная в данной работе, была обучена на небольшом объёме данных и их увеличение приведёт более точным прогнозам. Так же для дальнейшего улучшения данной модели, необходимо использовать значения отражаемости по всем уровням, первичные радиометеорологические продукты лоаторов с двойной поляризацией, данные о радиальной скорости. Оптимизация обработки данных позволит обучить модель не только на одном метеорологическом лоаторе, но и на других. Так же есть возможность использования композитных изображений нескольких лоаторов одновременно. Подбор различных гиперпараметров и тестирование других сверточных нейронных сетей приведёт к росту точности и уменьшению функции ошибки. Для широкого использования на сети Росгидромета и Авиаметттелекома, необходимо интегрировать работу данной модели в имеющиеся пакеты приложений для обработки и визуализации данных метеорологических лоаторов.

Данная работа является первым шагом к дальнейшему развитию краткосрочного прогноза с помощью алгоритмов машинного обучения.

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Методические указания по использованию информации доплеровского метеорологического радиолокатора ДМРЛ-С в синоптической практике. М. Третья редакция, 2019 г.
2. Отчёт по теме 9.15 “Методическое руководство метеорологическими радиолокационными наблюдениями на ГНС Росгидромета” Плана оперативно-производственных работ НИУ Росгидромета за 1-й квартал 2022г. Аналитический отчёт о результатах р/л наблюдений на сети ДМРЛ Росгидромета в 4 квартале 2021г. ФГБУ “ЦАО” 2022 г.
3. Яндекс.Погода, «Метеум». <https://yandex.ru/pogoda/meteum>
4. Deep Learning by Ian Goodfellow, Yoshua Bengio and Aaron Courville published by MIT Press, 2016
5. Stanford University’s Course — CS231n: Convolutional Neural Network for Visual Recognition by Prof. Fei-Fei Li, Justin Johnson, Serena Yeung
6. <https://datascience.stackexchange.com/questions/14349/difference-of-activation-functions-in-neural-networks-in-general>
7. [https://www.codementor.io/james\\_aka\\_yale/convolutional-neural-networks-the-biologically-inspired-model-iq6s48zms](https://www.codementor.io/james_aka_yale/convolutional-neural-networks-the-biologically-inspired-model-iq6s48zms)
8. <https://searchenterpriseai.techtarget.com/definition/convolutional-neural-network>
9. Convolutional neural networks: an overview and application in radiology.  
<https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9>
10. <http://library.eltech.ru/files/vkr/2017/magistri/1307/2017BKP130724ГРЕЧК О.pdf>.
11. [https://nure.ua/wp-content/uploads/2019/Scinetific\\_editions/radio\\_engineering\\_197/4.pdf](https://nure.ua/wp-content/uploads/2019/Scinetific_editions/radio_engineering_197/4.pdf).
12. <https://habr.com/ru/company/yandex/blog/317626/>.

