



ТОМСКИЙ  
ПОЛИТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ



# МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ СНИМКОВ С БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

К. А. Маслов: студент ОИТ ТПУ [kam20@tpu.ru](mailto:kam20@tpu.ru)

О. С. Токарева: доцент ОИТ ТПУ, к.т.н.

И. А. Керчев: с.н.с. ИМКЭС СО РАН, к.б.н.

Н. Г. Марков: профессор ОИТ ТПУ, д.т.н.

18 ноября  
2020

# Введение

Одним из основных негативных факторов, воздействующих на пихтовые леса ряда регионов Южной Сибири, Байкальского региона и Предуралья, является инвазивный короед – уссурийский полиграф (*Polygraphus proximus* Blandf.). Обширные очаги усыхания пихтарников, вызванные короедом, влекут за собой значимые экономические и экологические последствия



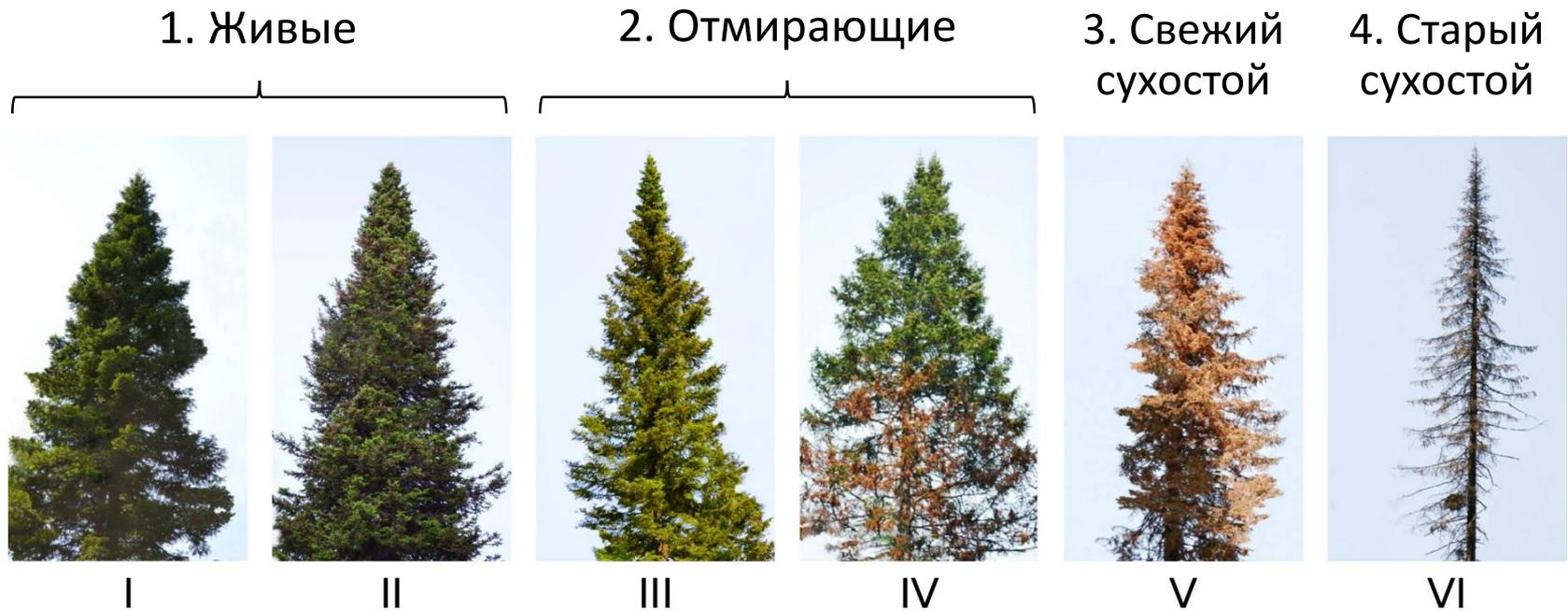
# Цель работы

Разработка алгоритмов машинного обучения для семантической сегментации деревьев пихты сибирской, поврежденных уссурийским полиграфом, на снимках с беспилотных летательных аппаратов

## Задачи

- Подготовить исходные данные
- Реализовать алгоритмы обучения
  - Текстурные признаки Харалика + градиентный бустинг
  - Полносверточные сети SegNet и U-Net
- Провести оценку эффективности разработанных алгоритмов

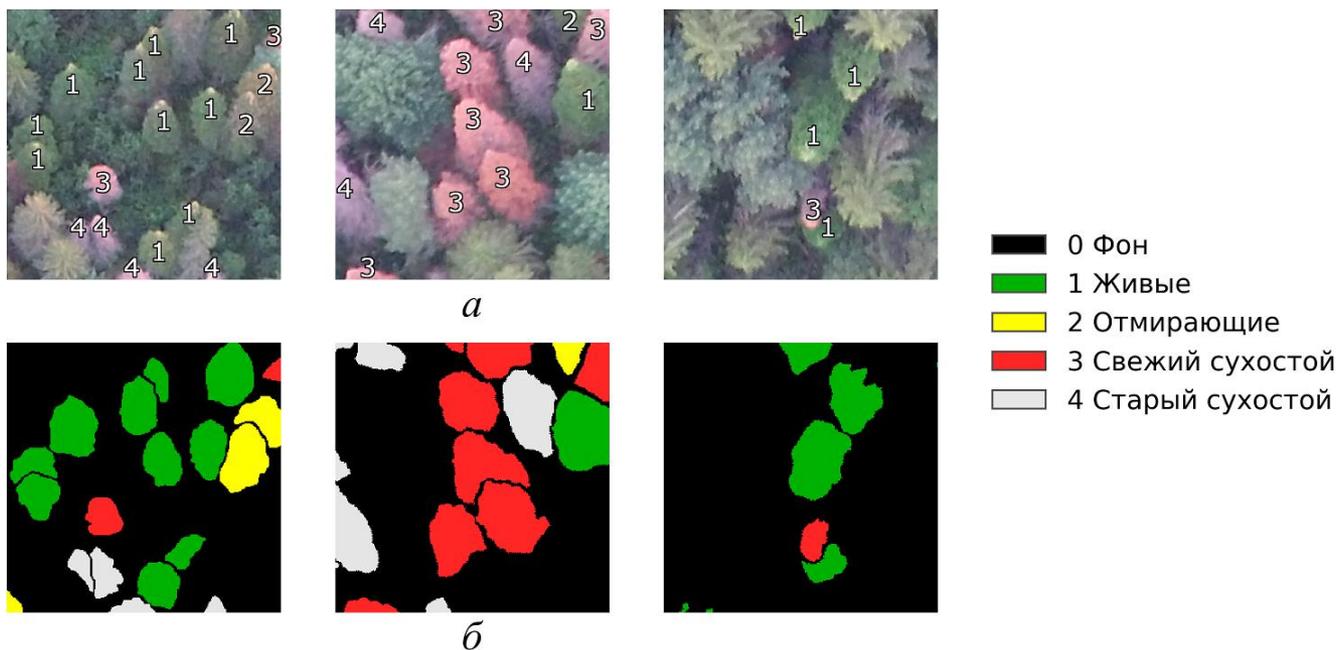
# Классы деревьев на снимках с БПЛА



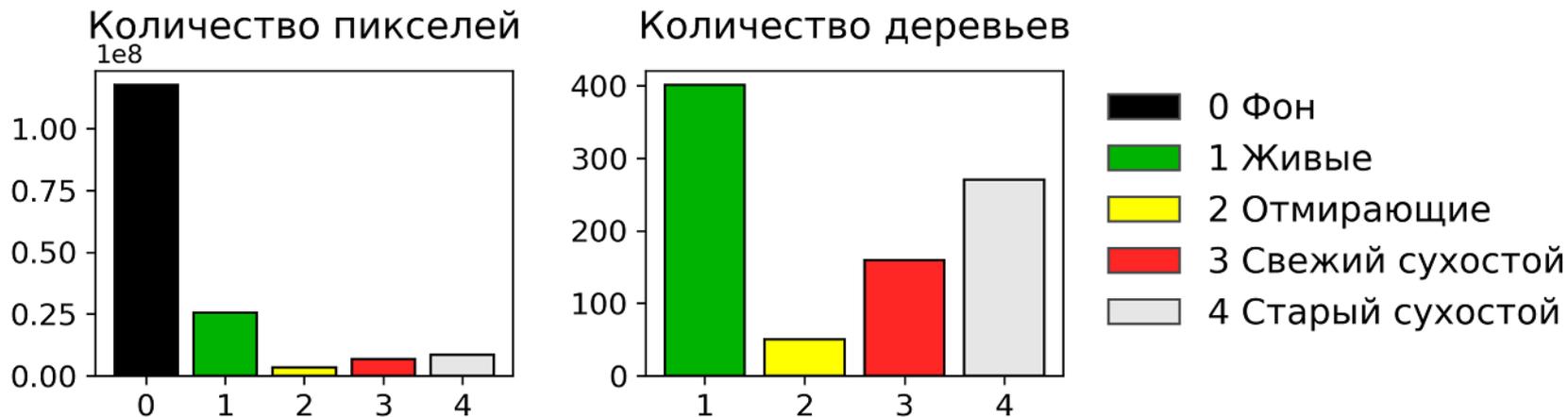
*Кривец С. А., Бисирова Э. М., Волкова Е. С., Дебков Н. М., Керчев И. А., Мельник М. А., Никифоров А. Н., Чернова Н. А. Технология мониторинга пихтовых лесов в зоне инвазии уссурийского полиграфа в Сибири. Методическое пособие. Томск: УМИУМ, 2018. 74 с.*

# Исходные данные

- Снимки получены с дрона DJI Phantom 3 Standard
- Съёмка проводилась в видимом диапазоне электромагнитного спектра (RGB)
- Период съёмки – с 7 по 28 августа 2017 г.
- Высота съёмки – около 350 м
- Пространственное разрешение около 0,1 м



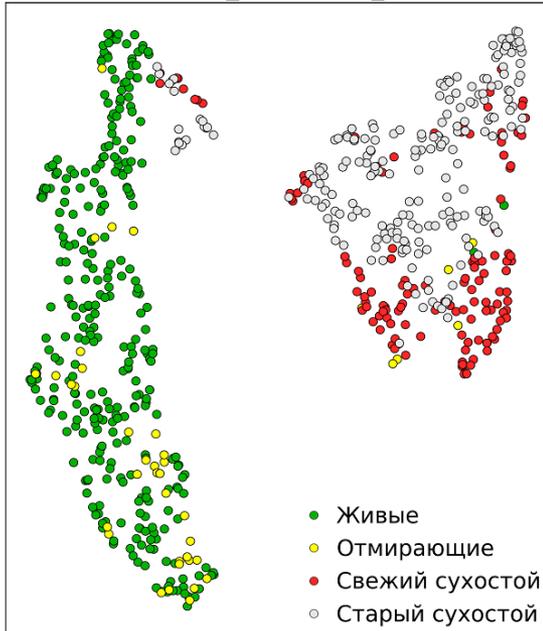
# Соотношения классов в исходных данных



- Значительное преобладание пикселей фона
- Наименее представлены деревья класса «Отмирающие»

# UMAP-вложения и диаграмма силуэтов

UMAP-вложение для гистограмм сегментов ( $n_n=10$ ,  $\text{min\_dist}=0.2$ )



UMAP-вложение для гистограмм сегментов ( $n_n=50$ ,  $\text{min\_dist}=1$ )

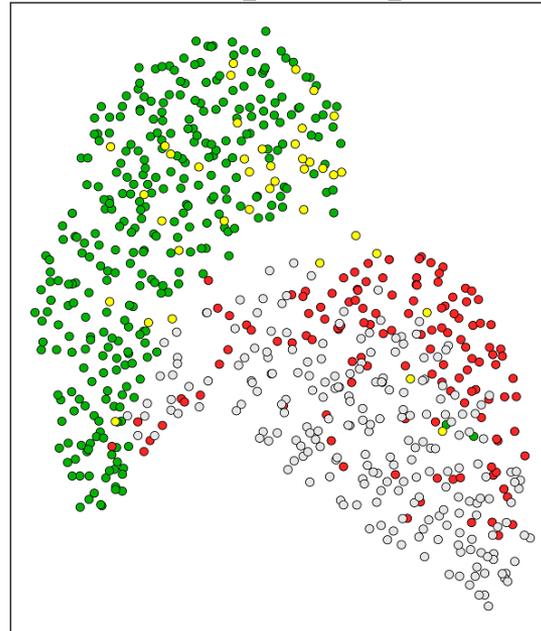
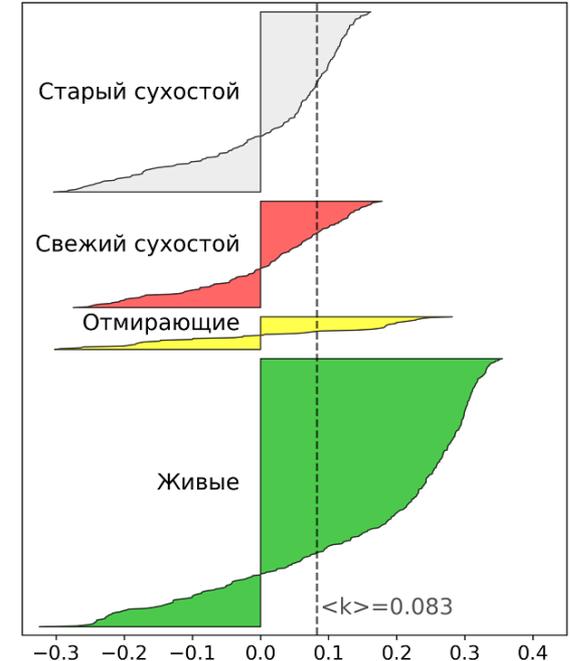


Диаграмма силуэтов



- Наблюдается хорошая разделимость классов сухостоя от остальных
- Наблюдается характерный непрерывный переход между классами
- Большой доле сегментов соответствуют отрицательные значения коэффициента силуэтов

# Текстурные признаки Харалика (1)

Основаны на GLCM-матрицах:

$$GLCM_{ij}(\theta_d, 0^\circ) = |\{(k, l), (m, n) \in (L \times L) \times (L \times L) \mid k = m, |l - n| = \theta_d, I(k, l) = i, I(m, n) = j\}|,$$

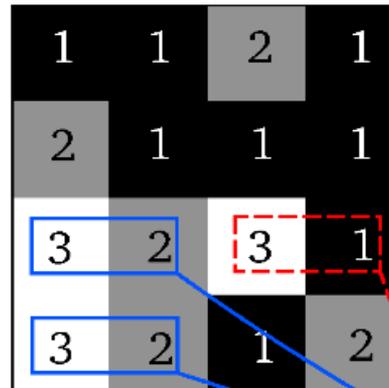
$$GLCM_{ij}(\theta_d, 45^\circ) = |\{(k, l), (m, n) \in (L \times L) \times (L \times L) \mid (k - m = \theta_d, l - n = -\theta_d) \text{ or } (k - m = -\theta_d, l - n = \theta_d), I(k, l) = i, I(m, n) = j\}|,$$

$$GLCM_{ij}(\theta_d, 90^\circ) = \dots,$$

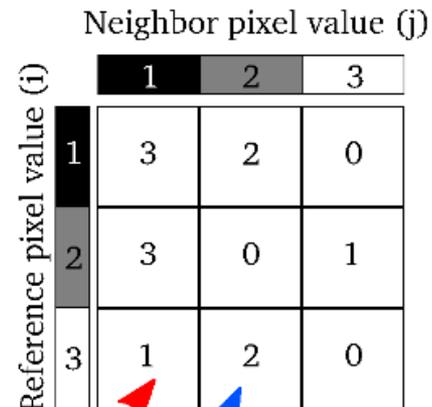
$$GLCM_{ij}(\theta_d, 135^\circ) = \dots,$$

$$L = \{1, 2, \dots, \theta_{ps}\}$$

Image with numeric gray levels



Right neighbor GLCM



# Текстурные признаки Харалика (2)

- Contrast 
$$\sum_{i,j=0}^{levels-1} GLCM_{ij} (i-j)^2$$

- Dissimilarity 
$$\sum_{i,j=0}^{levels-1} GLCM_{ij} |i-j|$$

- Homogeneity 
$$\sum_{i,j=0}^{levels-1} \frac{GLCM_{ij}}{1+(i-j)^2}$$

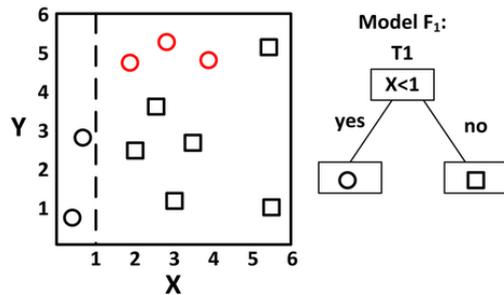
- ASM 
$$\sum_{i,j=0}^{levels-1} GLCM_{ij}^2$$

- Energy 
$$\sqrt{ASM}$$

- Correlation 
$$\sum_{i,j=0}^{levels-1} GLCM_{ij} \left[ \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \right]$$

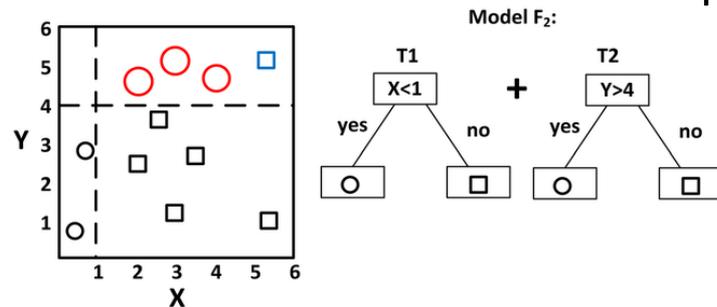
# Градиентный бустинг

Iteration 1

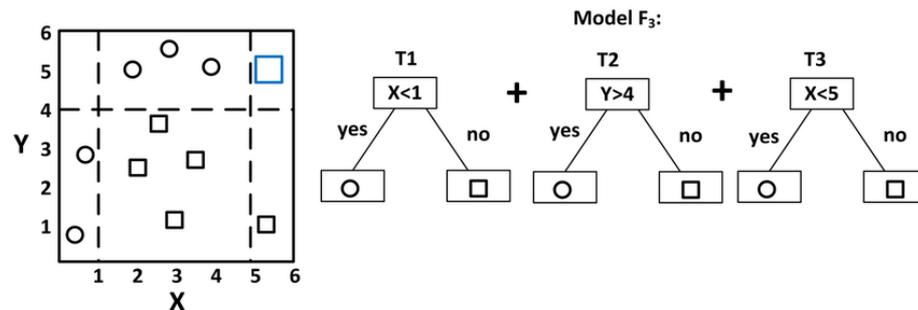


– техника построения ансамблей моделей (обычно, неглубоких деревьев решений), в которой предсказатели построены последовательно, причем, каждый следующий предсказатель пытается исправить ошибку предыдущих.

Iteration 2

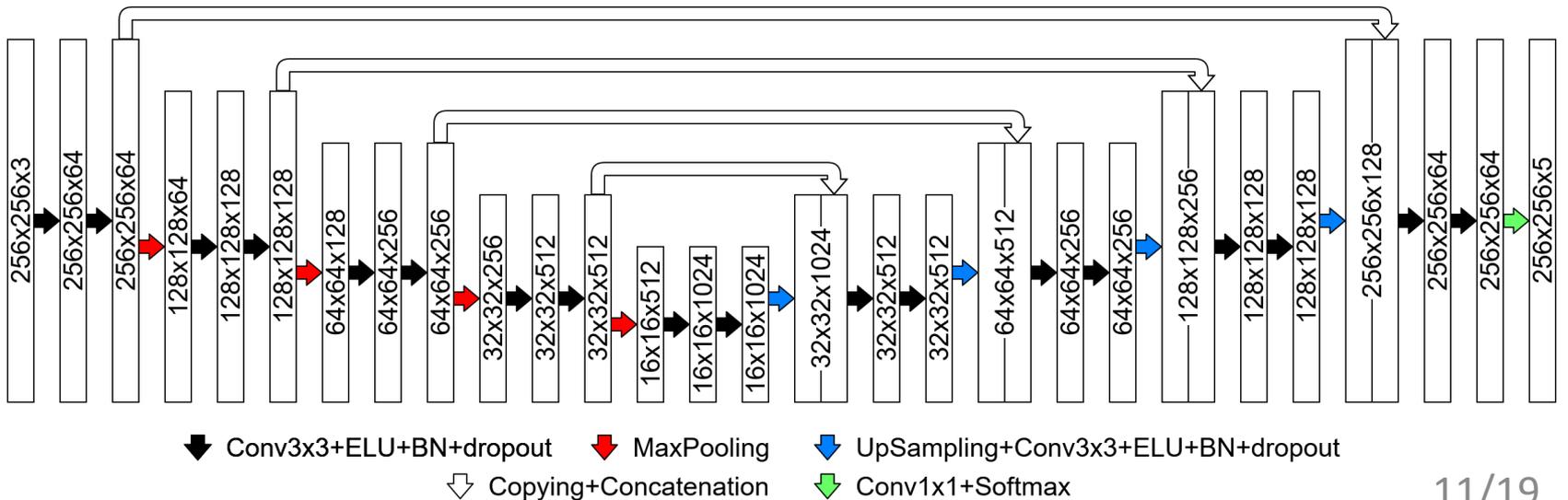
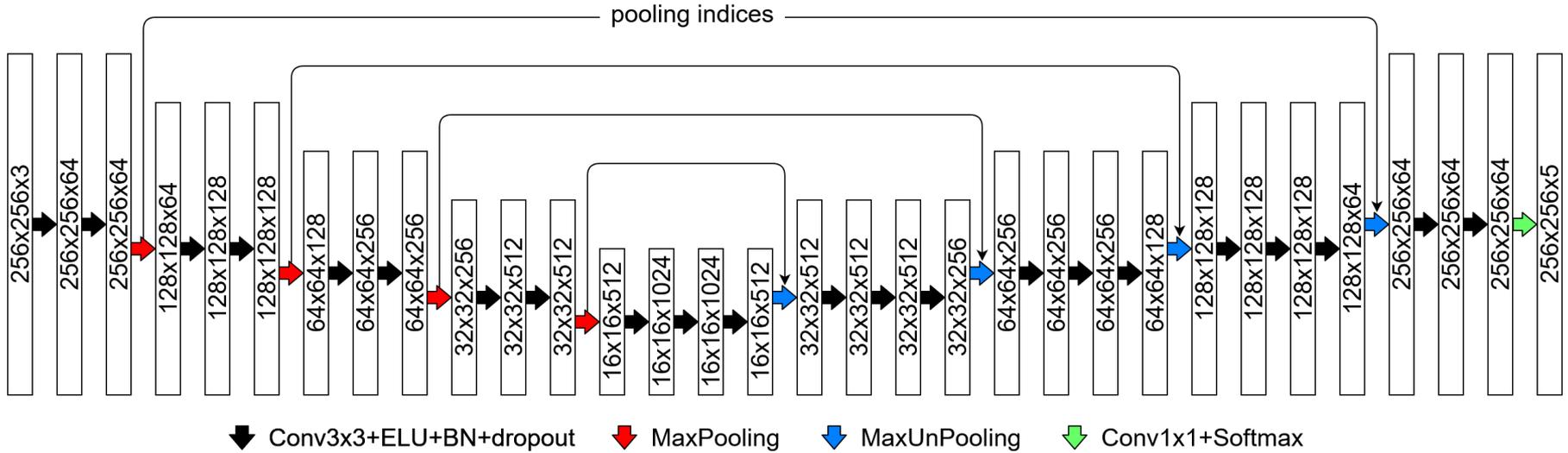


Iteration 3



Zhang Z. et al. Exploring the clinical features of narcolepsy type 1 versus narcolepsy type 2 from European Narcolepsy Network database with machine learning, *Scientific Reports*, 2018. Vol. 8, No. 1. DOI:10.1038/s41598-018-28840

# Полносверточные сети SegNet и U-Net

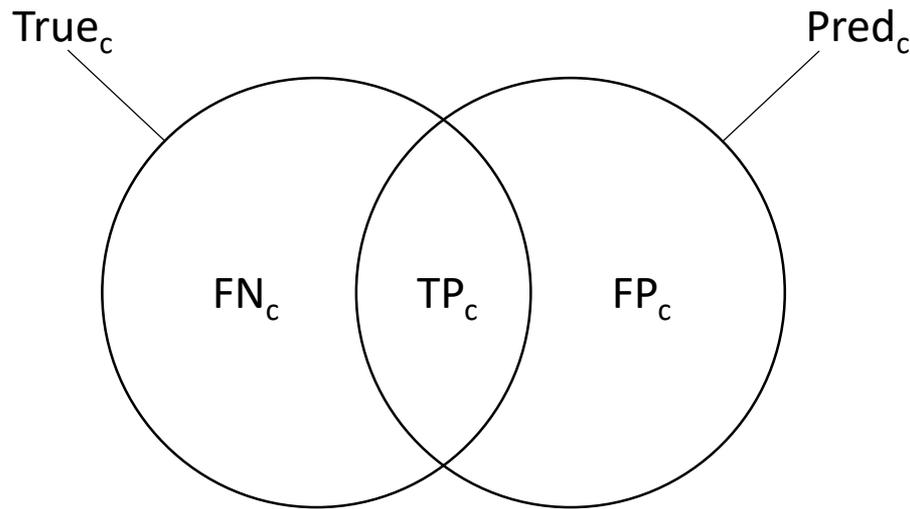


# Подготовка данных

- Для минимизации разницы в условиях освещенности различных снимков была проведена процедура выравнивания гистограмм
- Для обучения градиентного бустинга из обучающих участков извлекались пиксели с шагом 10, в качестве признаков брались спектральные характеристики пикселя и рассчитанные текстурные признаки Харалика в его окрестности
- Для обучения нейронных сетей путем поворотов и нарезки снимков было получено 2469 обучающих и 360 валидационных примеров
- С целью улучшения обобщающей способности нейронных сетей была использована аугментация «на лету», включающая случайные изменения масштаба, изменения яркости и отражения по вертикальной оси для обучающих примеров

# Целевая функция (1)

Целевой метрикой является усредненный коэффициент Жаккара (mIoU):



$$\text{IoU}_c = \frac{|\text{True}_c \cap \text{Pred}_c|}{|\text{True}_c \cup \text{Pred}_c|} = \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c + \text{FP}_c},$$

$$\text{mIoU} = \frac{1}{C} \sum_{c=1}^C \text{IoU}_c$$

## Целевая функция (2)

На этапе обучения нейронных сетей максимизировался mIoU:

$$\text{mIoU}(\mathbf{T}, \mathbf{P}) = \frac{1}{C} \sum_{c=1}^C \left( \frac{\sum_{i=1}^H \sum_{j=1}^W \text{LS}(\mathbf{T})_{ijc} \cdot \mathbf{P}_{ijc} + \theta_s}{\sum_{i=1}^H \sum_{j=1}^W [\text{LS}(\mathbf{T})_{ijc} + (1 - \text{LS}(\mathbf{T})_{ijc}) \cdot \mathbf{P}_{ijc}] + \theta_s} \right) \rightarrow \max,$$

$$\text{LS}(\mathbf{T}) = (1 - \theta_{ls}) \cdot \mathbf{T} + \frac{\theta_{ls}}{C},$$

где  $\mathbf{T}$  – тензор с эталонным выходом,  $\mathbf{P}$  – действительный выход нейронной сети,  $H$  – высота изображения,  $W$  – ширина изображения,  $C$  – количество классов,  $\theta_s$  – сглаживающий параметр, устраняющий большие по модулю значения градиента целевой функции и точку сингулярности,  $\theta_{ls}$  – параметр сглаживания меток

# Гиперпараметры алгоритмов

Гиперпараметры градиентного бустинга:

$\theta_{ne}$  – максимальное количество деревьев в ансамбле,

$\theta_{md}$  – максимальная глубина дерева решений,

$\theta_{lr}$  – скорость обучения,

$\theta_{ps}$  – размер окна,

$\theta_d$  – расстояние между пикселями

Гиперпараметры SegNet/U-Net:

$\theta_{lr}$  – скорость обучения,

$\theta_d$  – коэффициент дропаута,

$\theta_s$  – сглаживание целевой функции,

$\theta_{ls}$  – сглаживание меток,

$\theta_z$  – изменение масштаба

фрагментов,

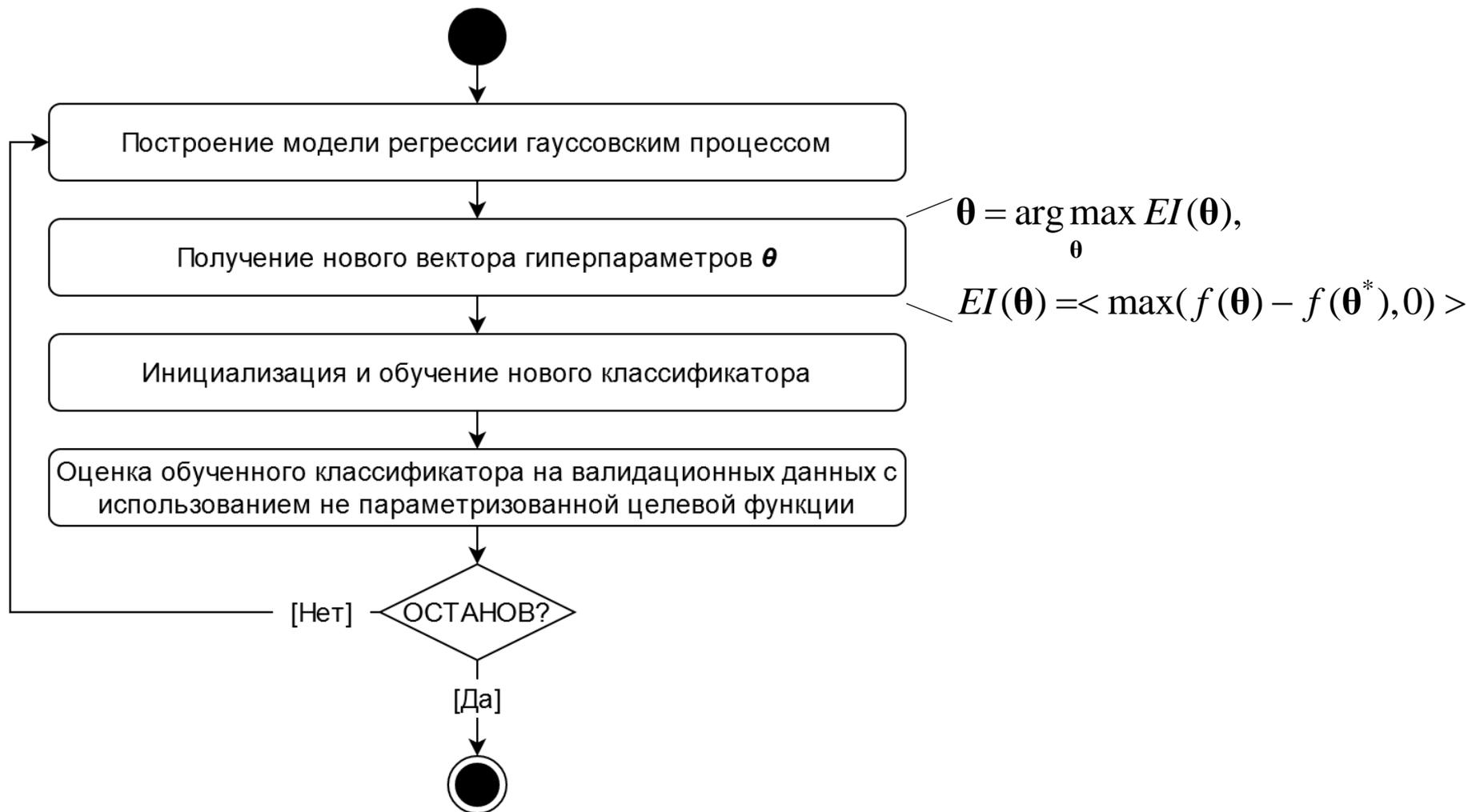
$\theta_b$  – изменение яркости фрагментов,

$\theta_{pf}$  – уменьшение скорости обучения при выходе «на плато»,

$\theta_{pp}$  – количество эпох для

подтверждения выхода «на плато»

# Подбор гиперпараметров



# Результат сегментации тестового участка

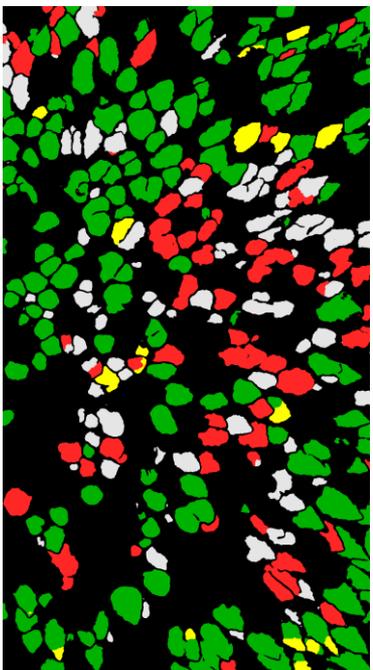
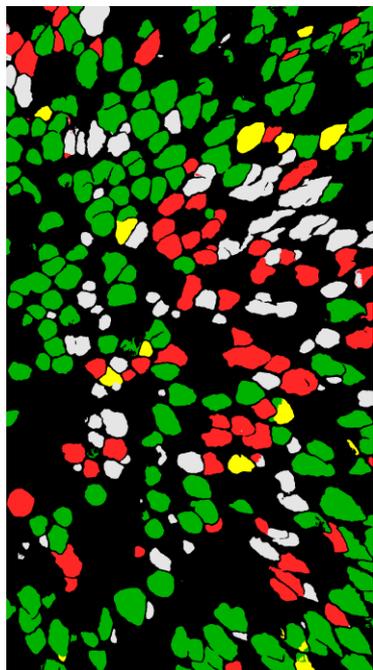
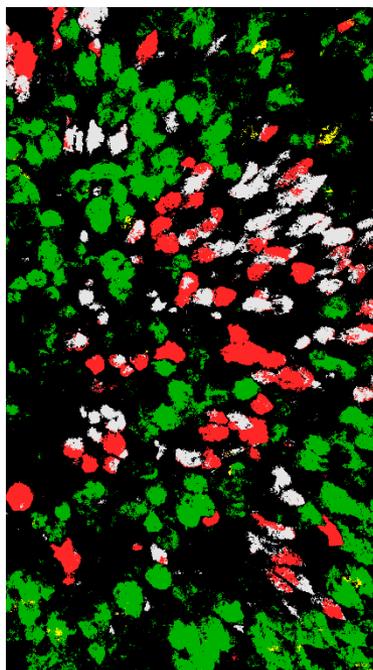
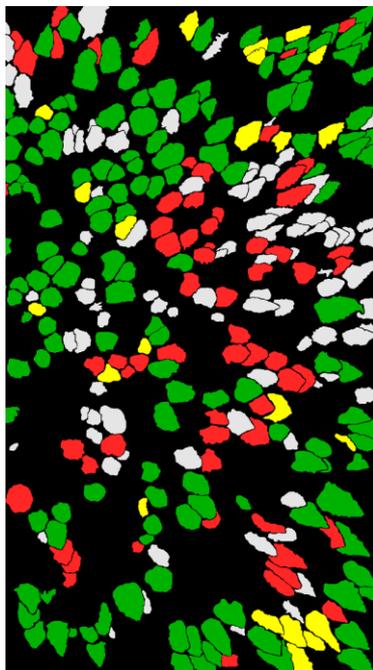
Участок

Эталон

Градиентный бустинг

SegNet

U-Net



■ Фон    ■ Живые    ■ Отмирающие    ■ Свежий сухостой    ■ Старый сухостой

# Метрики качества классификации

Алгоритм	IoU					mIoU
	Фон	Живые	Отмирающие	Свежий сухостой	Старый сухостой	
Градиентный бустинг	0,76	0,53	0,05	0,62	0,50	0,49
SegNet	0,85	0,74	0,38	0,75	0,65	0,67
U-Net	0,86	0,76	0,42	0,75	0,68	0,69

# Заключение

- Разработано три алгоритма для решения задачи сегментации поврежденной уссурийским полиграфом пихты на снимках с беспилотных летательных аппаратов
- Среди разработанных алгоритмов лучшую эффективность показала полносверточная сеть U-Net. Данная модель успешно классифицирует пиксели классов «Фон», «Живые», «Свежий сухостой» и «Старый сухостой» ( $IoU > 0,5$ )
- Эффективность предложенных подходов может быть улучшена путем увеличения количества данных для обучения и привлечения качественно новых данных (например, снимков в инфракрасном диапазоне электромагнитного спектра, карт глубины и т.д.)
- В связи с непрерывностью процесса ухудшения состояния древостоя, возможно, имеет смысл рассматривать задачу класса нечеткой классификации или регрессии