

Анализ изображений
в образовательном
тестировании с
помощью
машинного обучения
(на примере инструмента
измерения креативности)

*ПОДГОТОВИЛИ СОТРУДНИКИ
ЦЕНТРА ПСИХОМЕТРИКИ И ИЗМЕРЕНИЙ
В ОБРАЗОВАНИИ
ИНСТИТУТА ОБРАЗОВАНИЯ НИУ ВШЭ:*

*ТАРАСОВ СЕРГЕЙ
ГЕЛЬВЕР ЕВГЕНИЙ
ГРАЧЕВА ДАРЬЯ
УГЛАНОВА ИРИНА
ВЫРВА ЕЛЕНА*

Что изучаем?

- Оценка сложных, метапредметных навыков является актуальной задачей системы образования.
- Для оценивания сложных навыков требуется прибегать не только к традиционным форматам тестирования, но и к более современным.
- Часто инновационные форматы предлагают среду тестирования, в которой требуется графическое решение тестового задания, что особенно актуально для заданий на креативность.

MACHINE LEARNING

A close-up, slightly angled shot of a vintage white typewriter. A sheet of white paper is fed into the carriage, and the words 'MACHINE LEARNING' are printed in large, bold, black capital letters. Below the title, three lines of Russian text are printed. The typewriter's internal mechanism, including the typebars and carriage, is visible through a semi-circular opening. The brand name 'TIPPAE' is partially visible on the right side of the machine. The foreground shows the top of several yellowed keys.

избежать субъективности,
вносимую экспертами

экономия временных и
денежных ресурсов на
работе экспертов

незамедлительная
обратная связь

TIPPAE



Цель

Продemonстрировать возможности машинного обучения для анализа изображений без привлечения экспертов в образовательном тестировании

Опыт применения машинного обучения при анализе изображений

- В исследовании Walt, Schönberger, 2014 описывается набор алгоритмов обработки изображений, реализованных на языке программирования Python с открытым исходным кодом.
- В исследовании Pei, Xing, & Lee, 2019 с помощью статистического анализа была выявлена взаимосвязь между характеристиками изображений и уровнями успеваемости учащихся при написании научных текстов.
- Графическое решение задач можно встретить в статьях по изучению навыков наложения швов хирургами (Frischknecht, Kasten, Hamstra, et al., 2013). Простой компьютерный алгоритм может извлекать переменные из цифровых изображений шва и быстро обеспечивать количественную итоговую оценку.

Исследования по измерению креативности

Torrance, 1981

Оригинальность оценивается как отличие от большинства изображений, созданных в релевантной выборке.

Ward, 1994

Сравнивается изображение, сделанное тестируемым, с конкретным изображением - референсом.

Отличающиеся характеристики нарисованных
инопланетных существ от животных Земли или от
людей:

- нетипичное расположение частей тела
- отсутствие двусторонней симметрии
- описание необычных способностей

(Kim, Vincent, & Goncalo, 2013)



Детальность -

относится к свойствам решения задачи, которое выражается в умении тщательно продумать решение проблемы и детально проработать его, создав целостный образ.

Оригинальность -

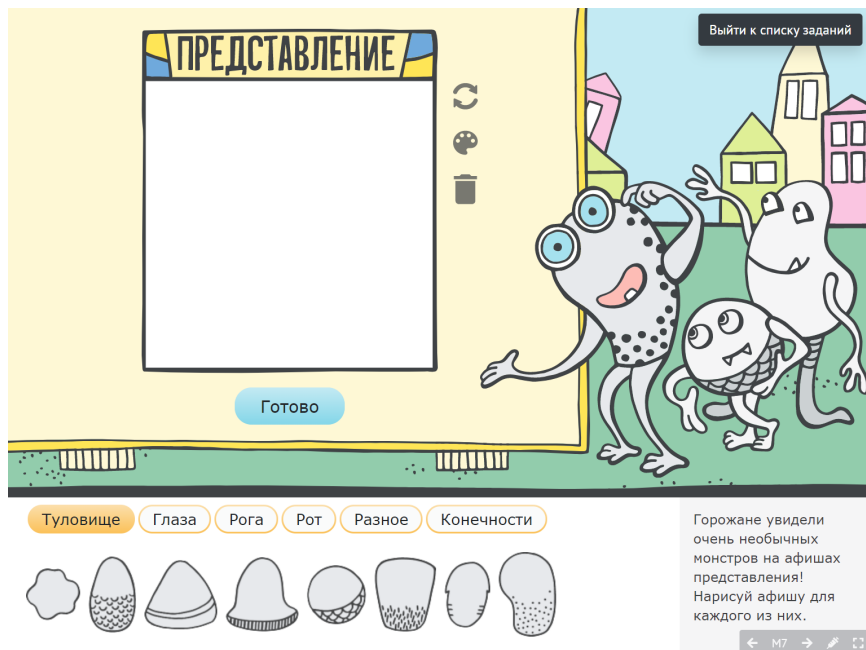
способность продуцировать новые идеи и решения задач, устанавливать неочевидные связи между идеями.

Креативность – способность представить и разработать принципиально новые подходы к решению проблем, ответы на вопросы, стоящие перед субъектом, или способы выражения идей для решения задач в повседневной жизни.

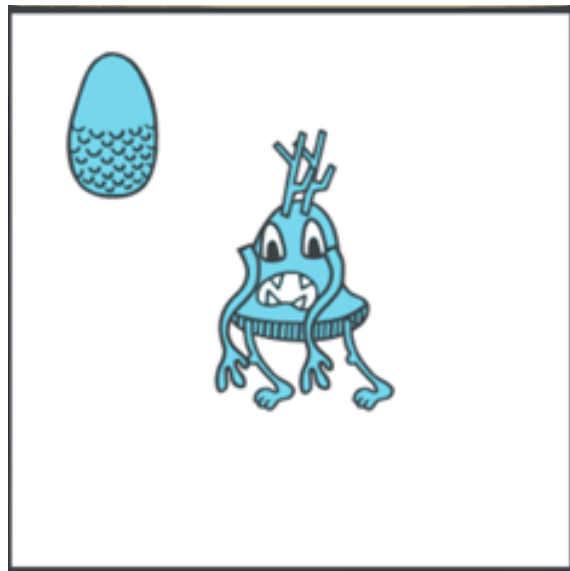
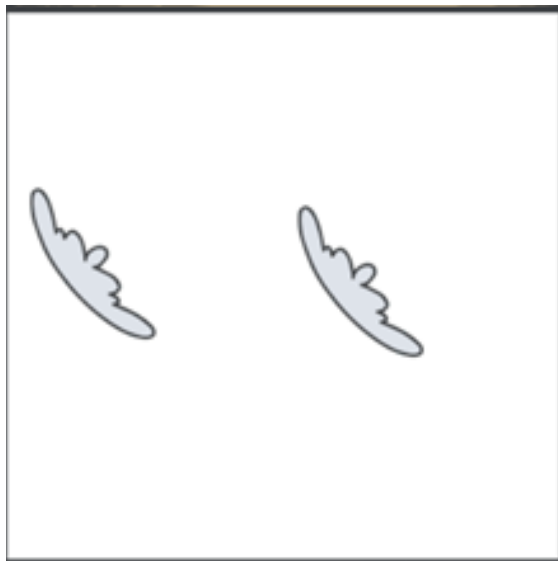
(Liu, Lin, Jian, Liou, 2012; Kupers, Lehmann-Wermser, McPherson, 2019)

Инструмент

- Для оценки навыков 21го века «4К» среди учеников 4 класса
- Задание "Монстр", оценивающее креативность
- Инструмент предьявляется тестируемому в компьютерной форме



ИНТЕРФЕЙС РАБОЧЕЙ ОБЛАСТИ ПОСТРОЕНИЯ МОНСТРА



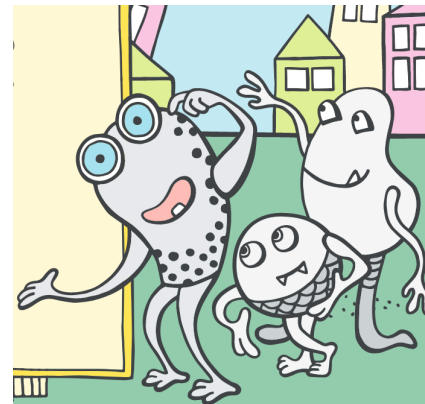
Примеры построенных монстров

Процедура сбора данных

Индикаторы оригинальности

Для оценки оригинальности использовались шесть индикаторов, оцениваемых дихотомически, где значение «1» присваивалось, если:

- расположение ног отлично от референса (не в нижней части тела);
- расположение рук отлично от референса (не в верхней части тела);
- расположение рта отлично от референса (не в нижней части туловища);
- расположение глаз отлично от референса (не в верхней части туловища);
- использовано больше 1 туловища;
- не наблюдается симметрия конечностей (рук и ног).



Индикаторы детальности

Для оценки детальности использовались два индикатора, оцениваемых дихотомически, где значение «1» присваивалось, если:

- был изменен цвет монстра
- было применено вращение к элементам монстра

И шесть индикаторов на общее количество элементов из каждой группы (туловища, разное, глаза, рот, рога, конечности).

Выборка

В исследовании принимали участие 1780 учащихся четвертых классов из Москвы, Калуги и Ярославля.

Так как каждый тестируемый строил по три монстра, то исходная база содержала 5340 изображения.

После удаления профилей с пропущенными значениями по всем индикаторам количество монстров в базе сократилось до 5286 (1762 профиля учащихся).



Структура анализа



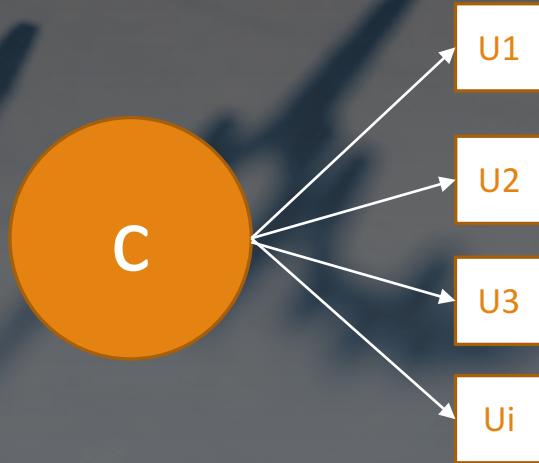
1 этап. Латентный классовой анализ - получение начальной разметки, которая будет использоваться для обучения нейросети.



2 этап. Машинное обучение - выбор архитектуры и обучение нейросети.

Методология LCA

Латентный классовой анализ позволяет по группе индикаторов выявить латентные классы монстров с различным уровнем выраженности оригинальности и детальности.



Этапы анализа

отдельно для детальности и оригинальности

1. Выявить оптимальное количество латентных классов

Построение моделей с разным количеством классов

Критерии выбора лучшей модели:

- Коэффициент энтропии (качество классификации) > 0.8
- Минимальные значения AIC и BIC
- Минимальная пропорция наблюдений в классах больше 8%
- Для парных сравнений тест Лу-Менделя-Рубина

2. Сравнить классы и дать им характеристику

Сравнение показателей отношения шансов принадлежности к классу для каждого индикатора

Результаты LCA

Оригинальность

Модель	Энтропия	Мин %	AIC	BIC	Статистика LMR
C=2	0.833	31	26075.438	26160.884	1509.007*
C=3	0.825	6	25911.146	26042.602	175.369*
C=4	0.573	6	25865.959	26043.425	58.217*

Выбрано двухклассовое решение

* - $p < 0.05$

Монстры в первом классе (69%)
«скорее не оригинальны», а
монстры во втором классе (31%)
«скорее оригинальны»

Индикатор	Отношение шансов
Расположение ног	1.56*
Расположение рук	3.43*
Расположение рта	26.27*
Расположение глаз	2.63*
Симметрия конечностей	1.58*

Результаты LCA

Детальность

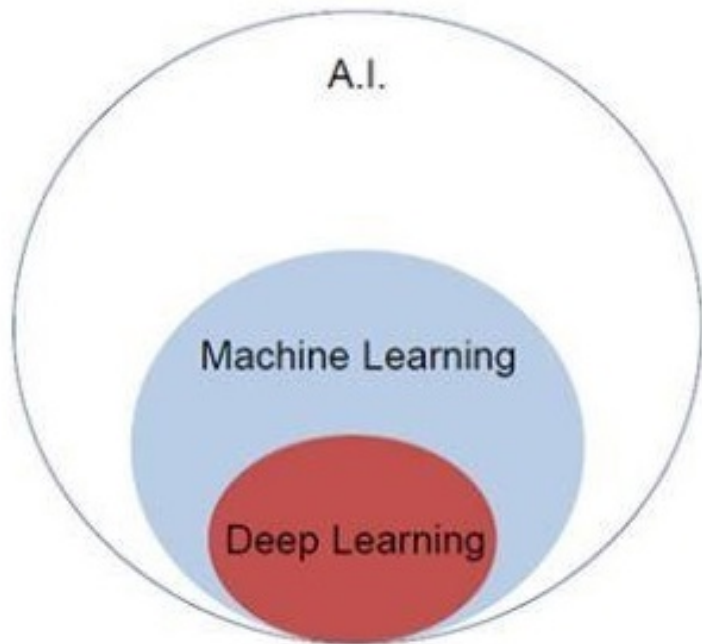
Модель	Энтропия	Мин %	AIC	BIC	Статистика LMR
C=2	0.816	16	63230.339	63419.769	2530.415*
C=3	0.660	16	62504.974	62792.386	749.509*
C=4	0.605	15	62329.210	62714.603	204.056*

Выбрано двухклассовое решение

Монстры в первом классе (16%) **«скорее детальные»**, а монстры во втором классе (84%) **«скорее не детальные»**

* - $p < 0.05$

Индикатор	Отношение шансов
Цвет >1	0.253*
Поворот >1	0.088*
Разное >1	0.263*
Разное >2	0.490*
Конечности >1	0.038*
Конечности >2	0.054*
Конечности >3	0.081*
Конечности >4	0.067*
Конечности >5	0.152*
Глаза >1	0.011*
Глаза >2	0.127*
Рот >1	0.068*
Рога > 1	0.19*
Рога > 2	0.158*



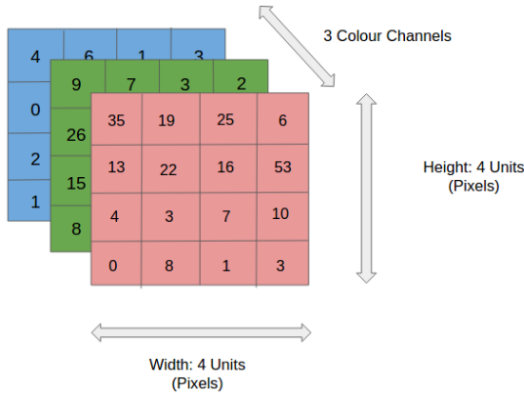
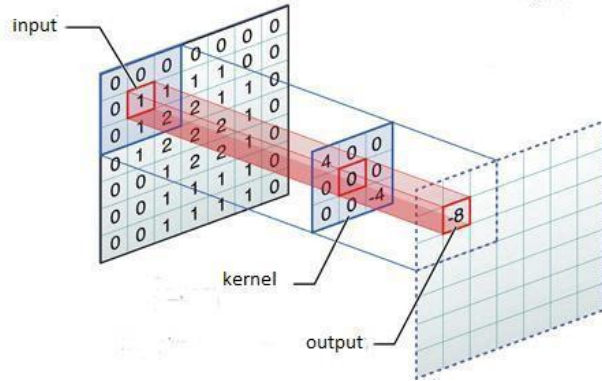
Методология анализа данных

Глубокое обучение (Deep Learning) - совокупность методов машинного обучения, основанных на обучении представлениям, а не специализированным алгоритмам под конкретные задачи.

Методология анализа данных

Сверточные нейронные сети (Convolutional Neural Networks)

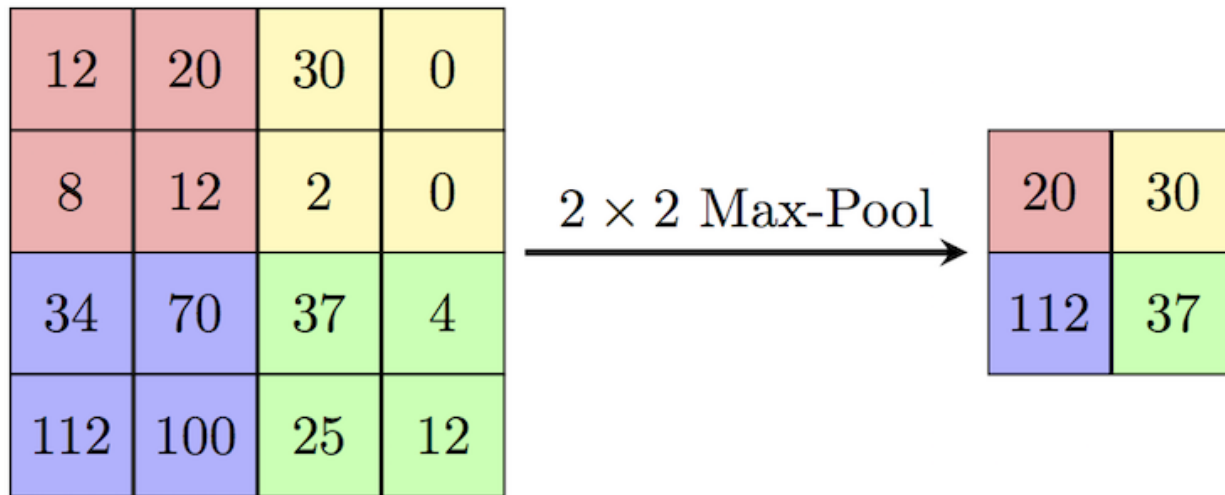
- *сверточный слой* - применение операции свертки к выходам с предыдущего слоя

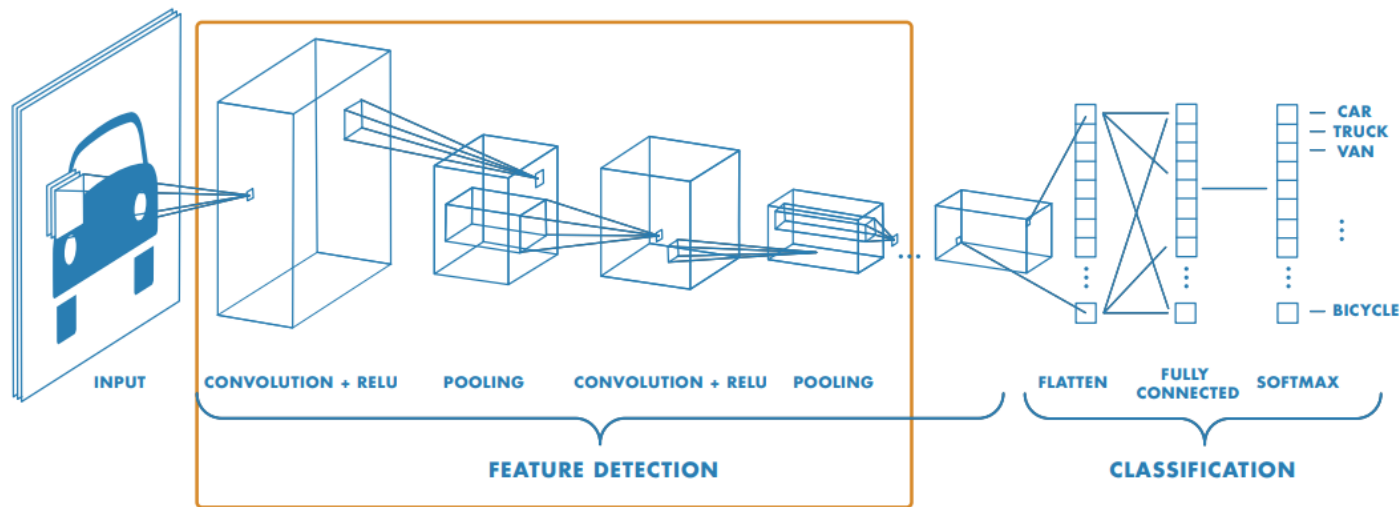


Сверточные нейронные сети (Convolutional Neural Networks)

- *субдискретизирующий слой* -
снижение размерности изображения

Методология анализа данных





Методология анализа данных

СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ (CONVOLUTIONAL NEURAL NETWORKS)

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

Методология анализа данных

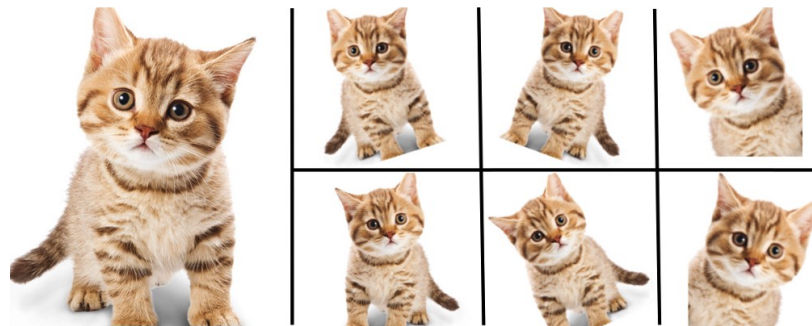
СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ (CONVOLUTIONAL NEURAL NETWORKS)

Аргументация данных (Data Augmentation)

Модель обучается путем минимизации эмпирического риска (увеличение точности), поэтому классы необходимо сбалансировать - нарастить недостающий класс.

1. Поворот изображения
2. Смена цвета
3. Отзеркаливание
4. Накладывание шума на пиксели

Методология анализа данных



Enlarge your Dataset

		Предсказанные значения	
		1	0
Истинные значения	1	606	135
	0	16	486

Результаты

ОРИГИНАЛЬНОСТЬ


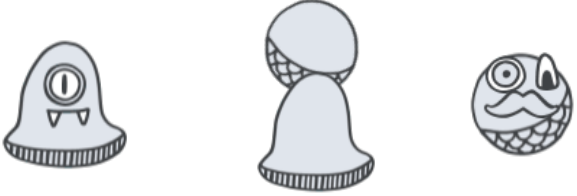
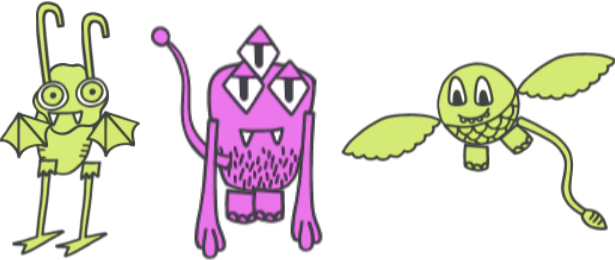

Точность (Accuracy) = $(606 + 486) / (606 + 486 + 135 + 16) = 0,878$



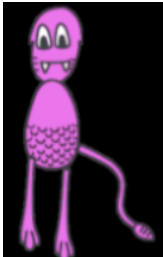

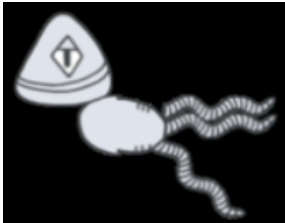

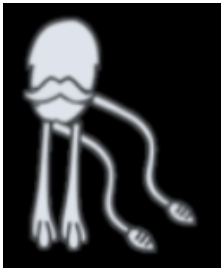
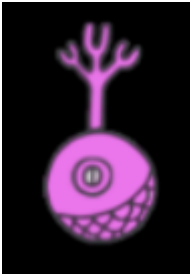




		Предсказанные значения	
		1	0
Истинные значения	1	603	135
	0	19	486

Результаты

ДЕТАЛЬНОСТЬ

Точность (Accuracy) = $(603 + 486) / (606 + 486 + 135 + 19) = 0,876$

	1	0
Оригинальность		
Детальность		

	False positive (если у сети 1, а у нас 0)	False negative (если у сети 0, а у нас 1)
Оригинальность	  	  
Детальность	  	  

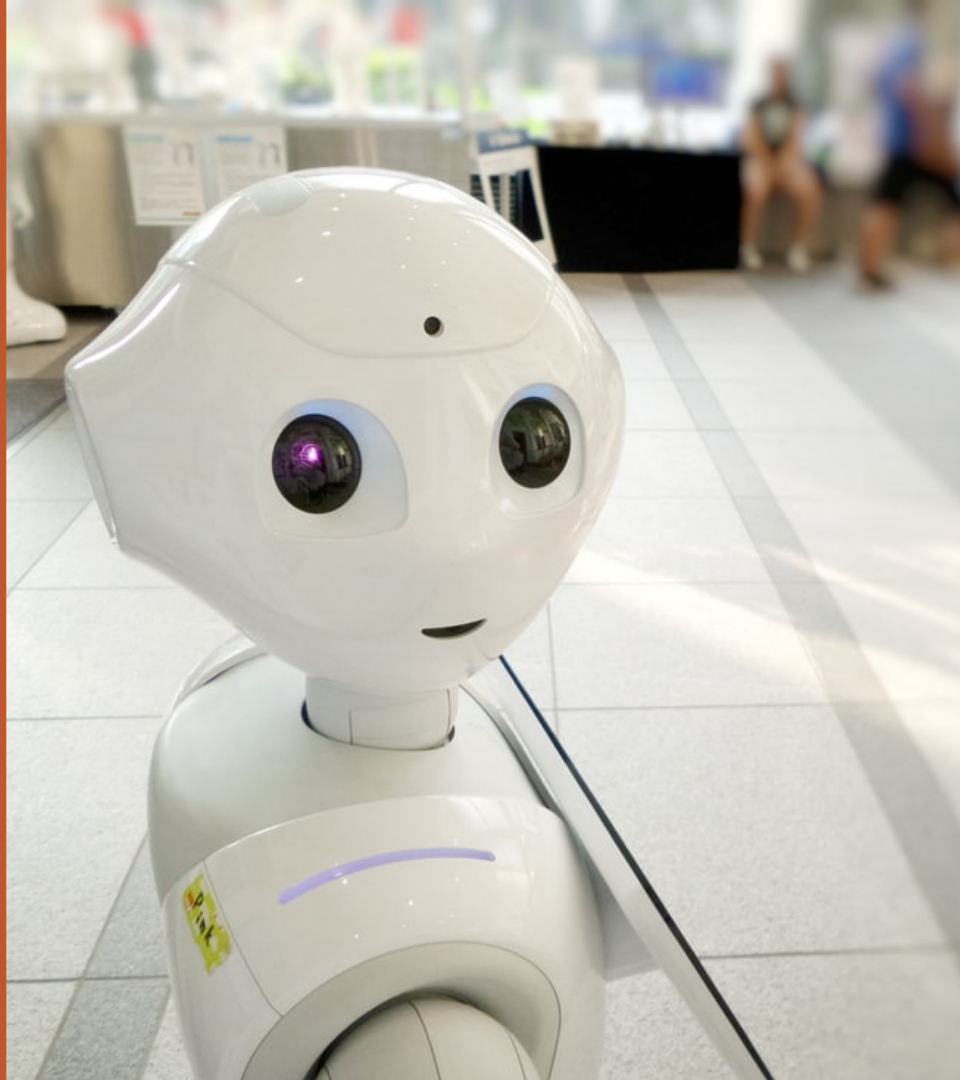
Дискуссия

Представлен подход к измерению креативности:

- 1) Разделение на группы по степени выраженности конструкта без привлечения экспертов – на основе теоретической модели с применением психометрических методов.
- 2) Автоматический анализ изображений – работа с изображением как с целостным объектом.

Подход решает проблемы, связанные с экспертной оценкой:

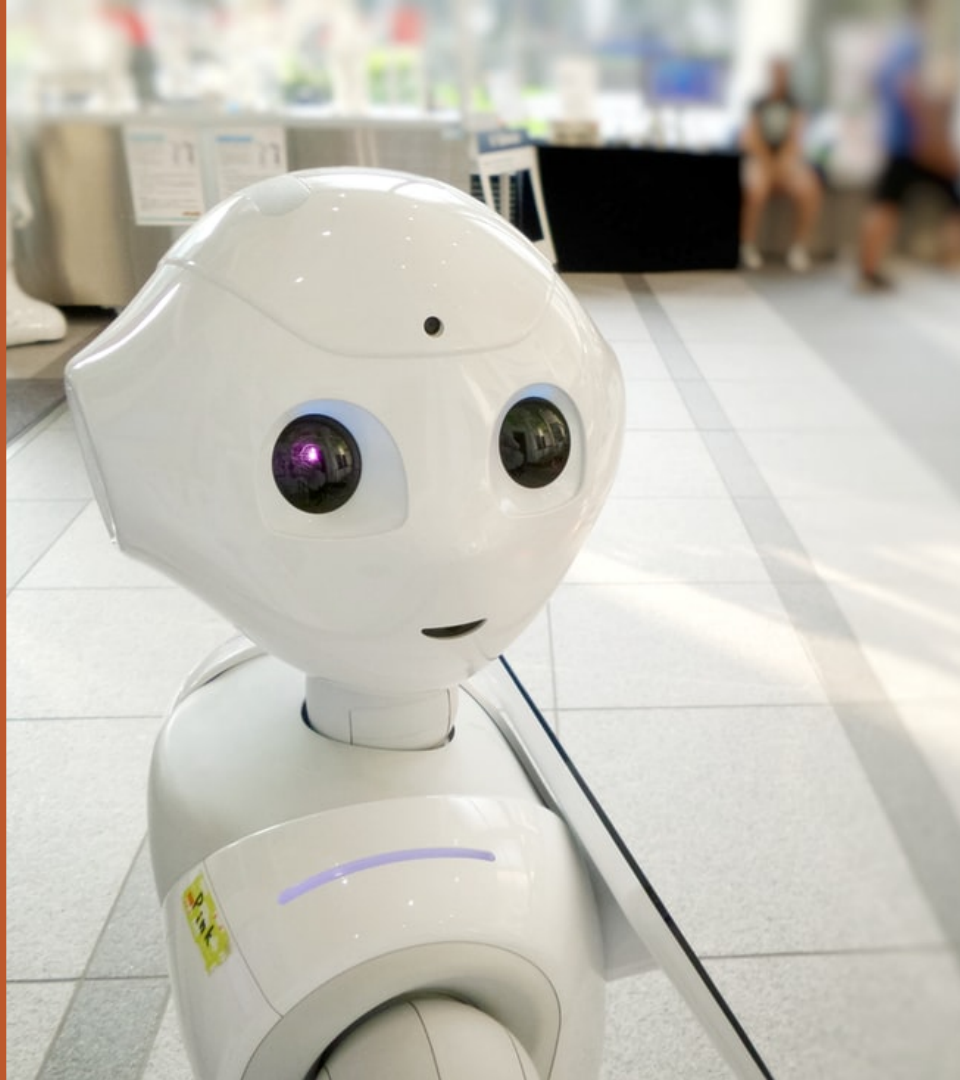
- ресурсозатратной работы экспертов;
- искажения (усталость, эффект ореола);
- проблема "смешивания" оригинальности и детальности



Дискуссия

Для успешного воспроизведения экспертной деятельности автоматизированными процессами машинного обучения необходимо обеспечить

- надлежащее качество данных
- качественную начальную разметку для обучения сети



Ограничения

- небольшой размеры выборки для обучения нейронной сети
- принадлежность трех монстров к одному ученику





Заключение

Новая, но крайне перспективная область исследований.

Развитие междисциплинарных исследований – наук о данных и наук об образовании.

Возможность ставить новые исследовательские вопросы и решать насущные практические задачи.



Нам нужны ваши вопросы!