Проект

Интерпретация модели ResNet9 на датасете SVHN

Выполнили:

Забродина Татьяна

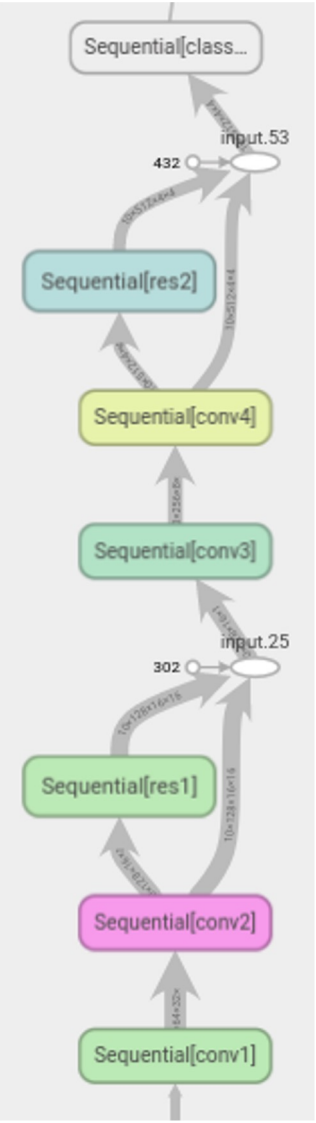
Шокоров Вячеслав

Про данные и модель

Датасет: SVHN (The Street View House Numbers)

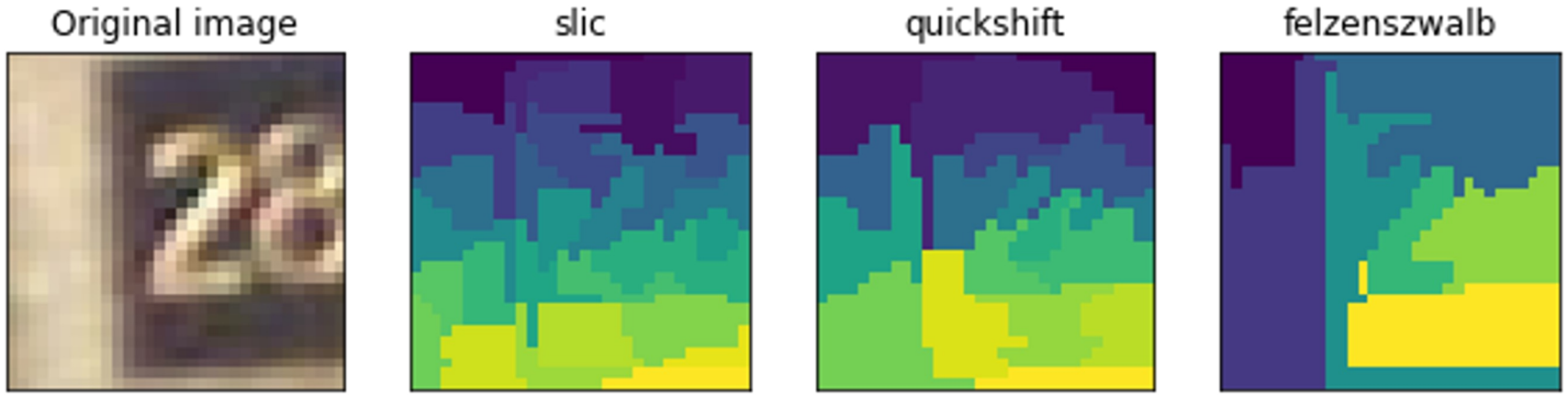
	Accuracy	ROC-AUC	Precision	Recall	F1_score
Train	0.967	0.998	0.966	0.966	0.966
Test	0.960	0.997	0.957	0.959	0.958

Модель: ResNet 9

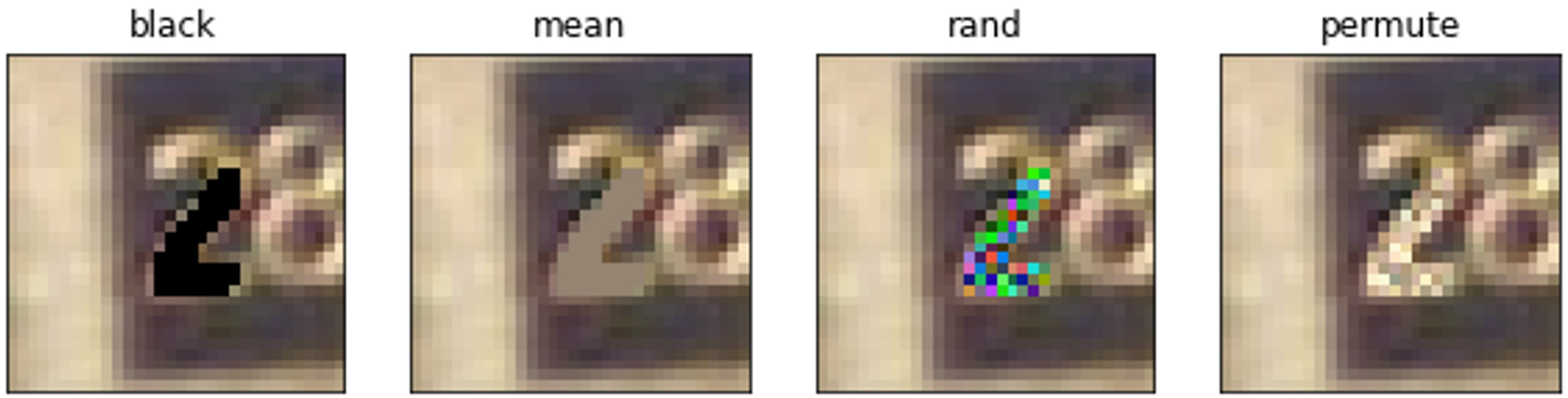


AIRI | Университет Сириус

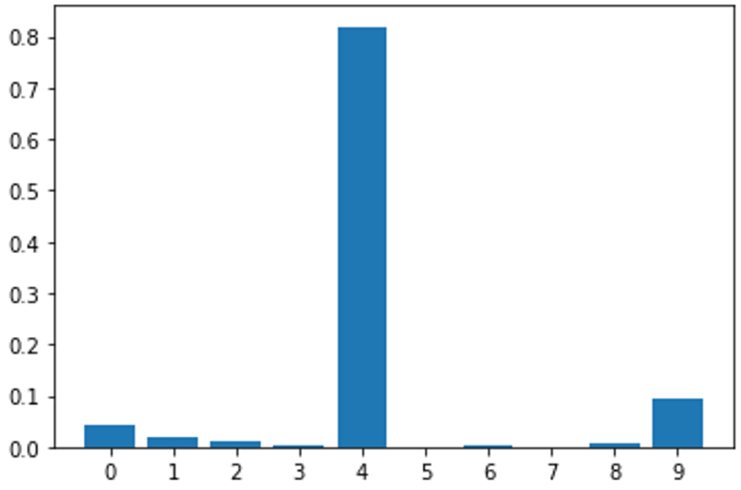
Разные варианты сегментации

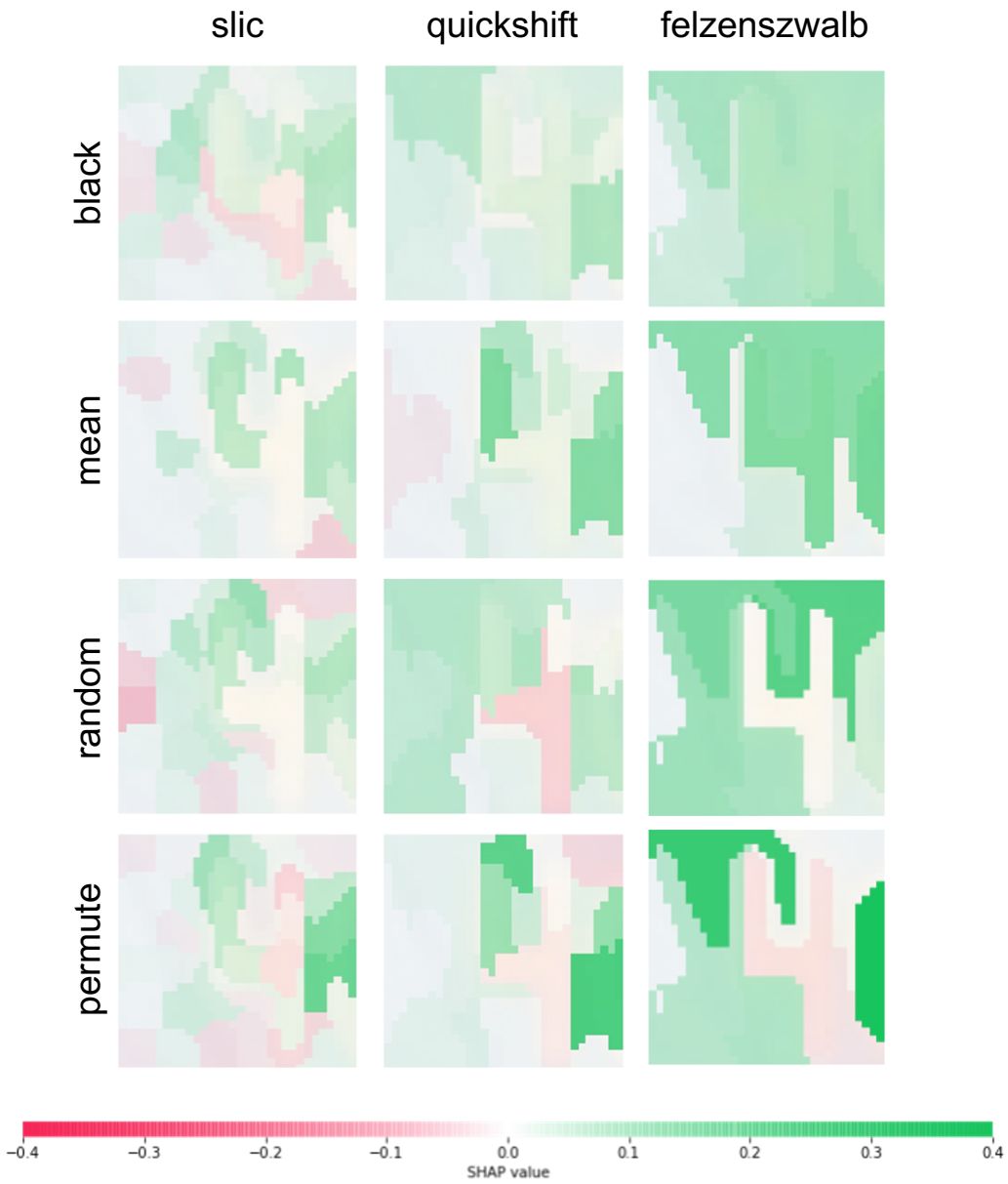


Разные варианты масок



Пример 10055

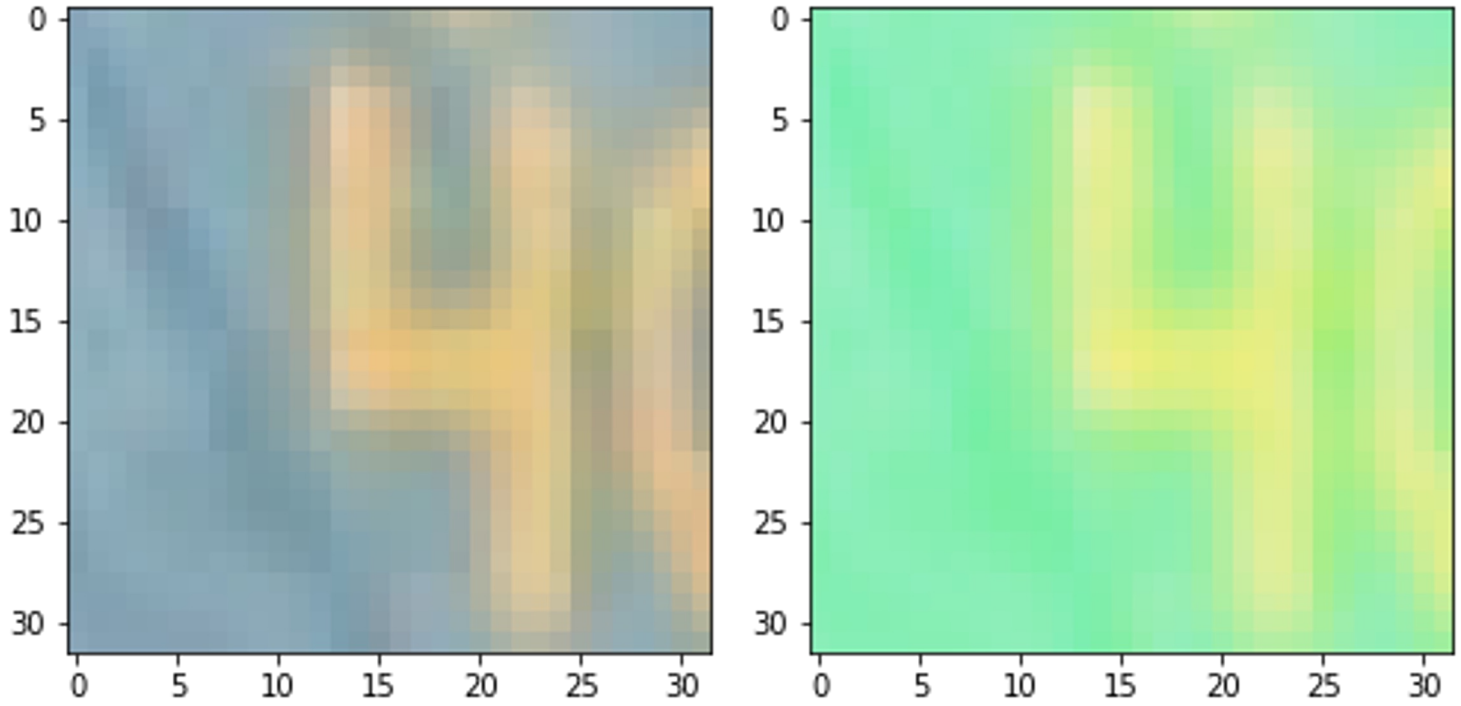




- При верном предсказании SHAP указывает, что модель не считает область цифры важной (в некоторых случаях указывает, что эта зона мешает предсказанию).
- При различной сегментации и различных методах заливки области SHAP дает противоречивые друг другу результаты. Это указывает на нестабильность SHAP-а.

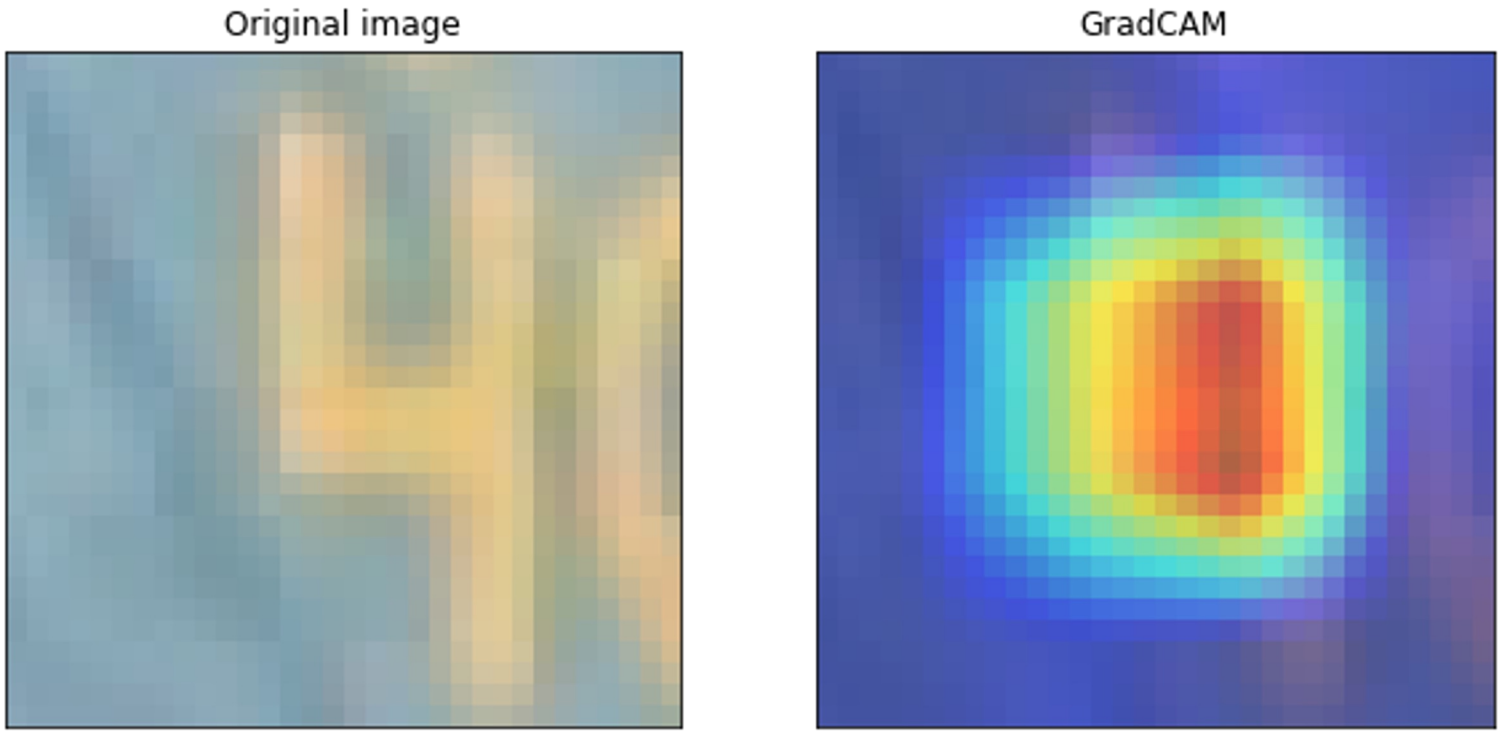
AIRI | Университет Сириус

Интерпретация LIME



- LIME не позволяет провести интерпретацию из-за того, что показывает, что модель смотрит на фон.

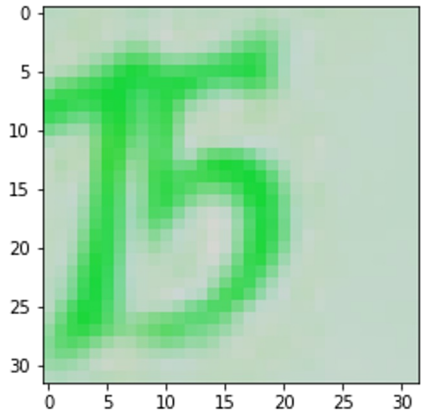
Интерпретация Grad-Cam



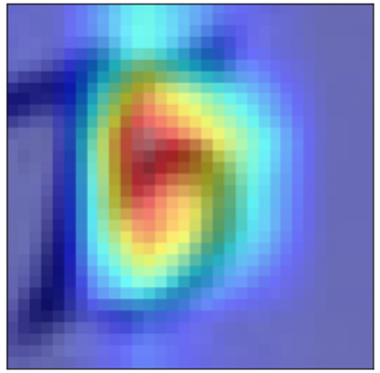
- Grad-Cam указывает на важность перемычки в цифре.

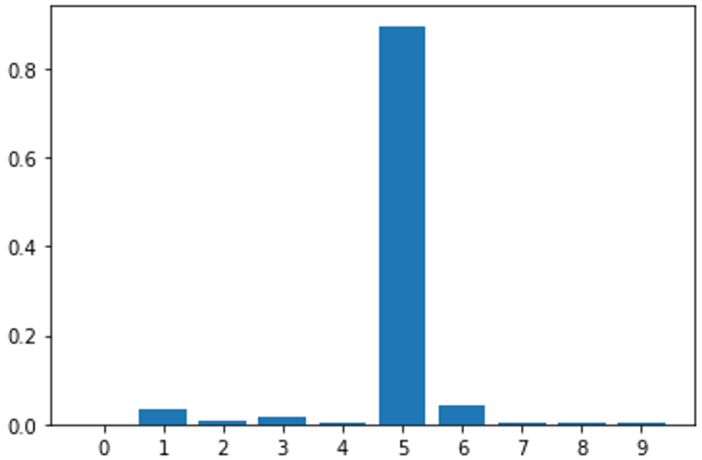
SHAP

LIME



Grad-Cam





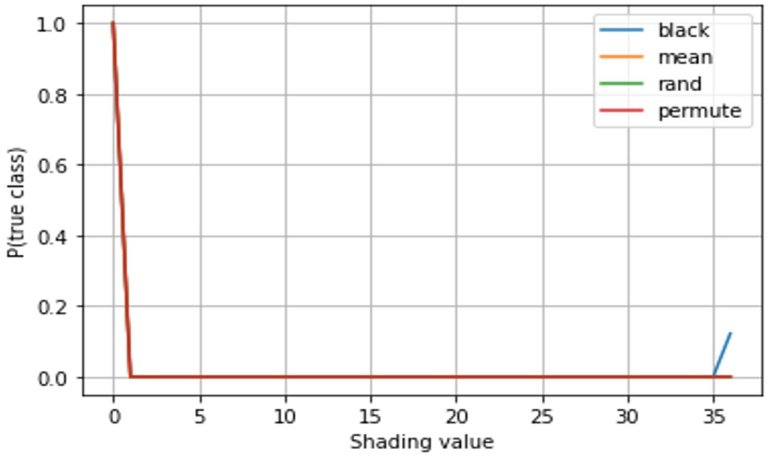
Комментарий:
SHAP – неудачная интерпретация
LIME – выделяет все цифры
Grad-Cam – указывает на особенности цифры, выбирая наиболее важные атрибуты.

AIRI | Университет Сириус

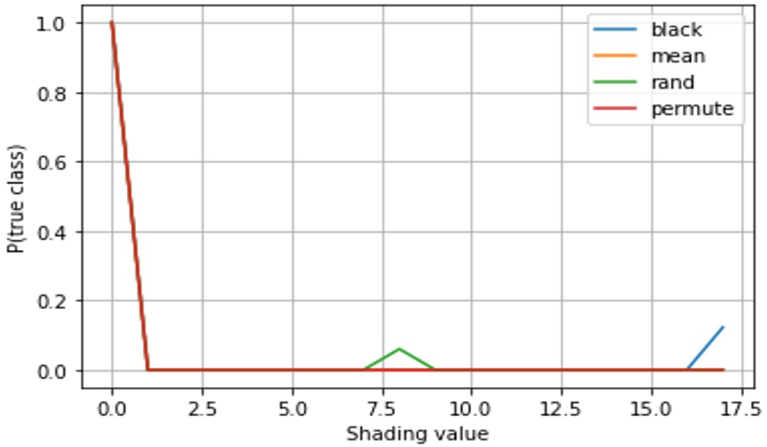
Последовательное затенение

Пример на slick

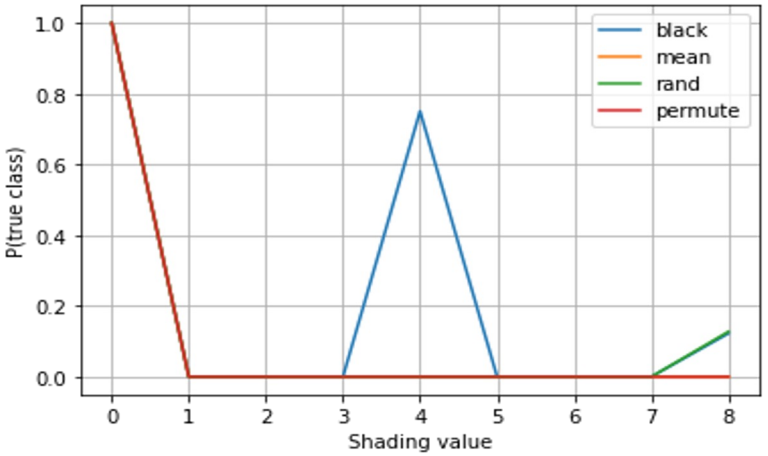
slick



quickshift



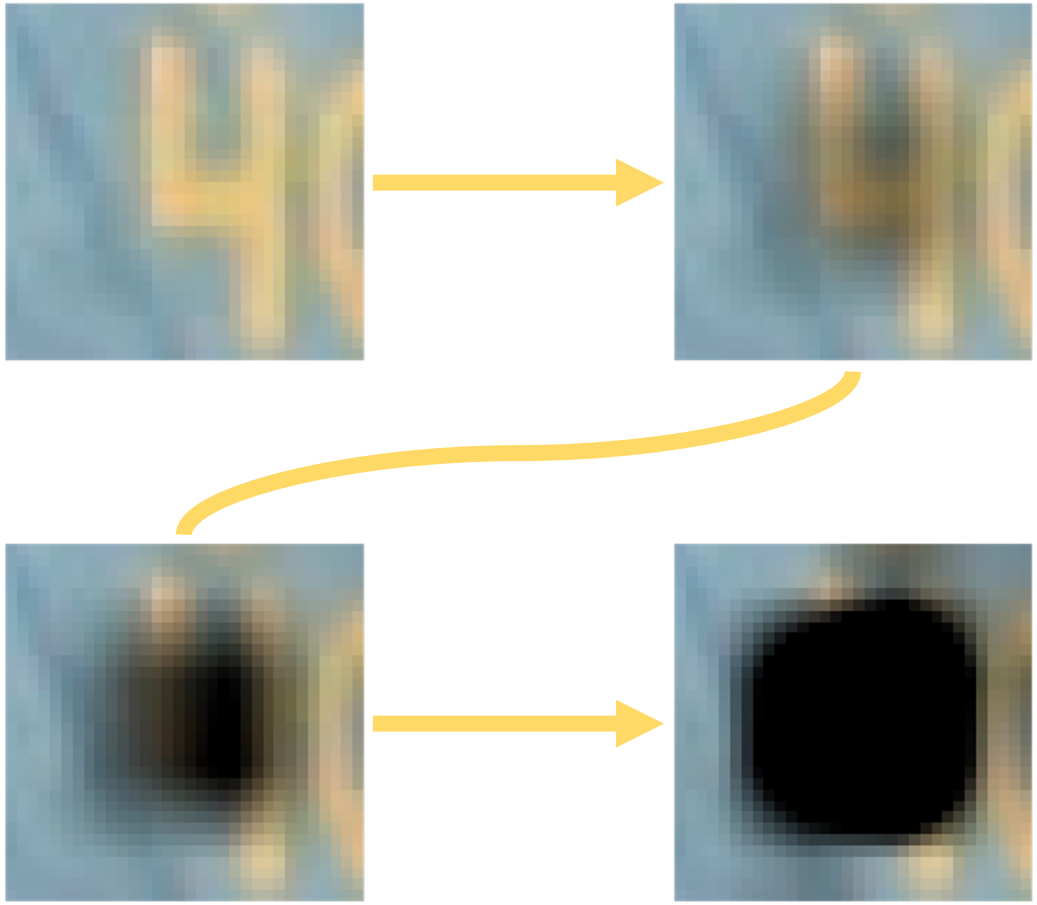
felzenszwalb



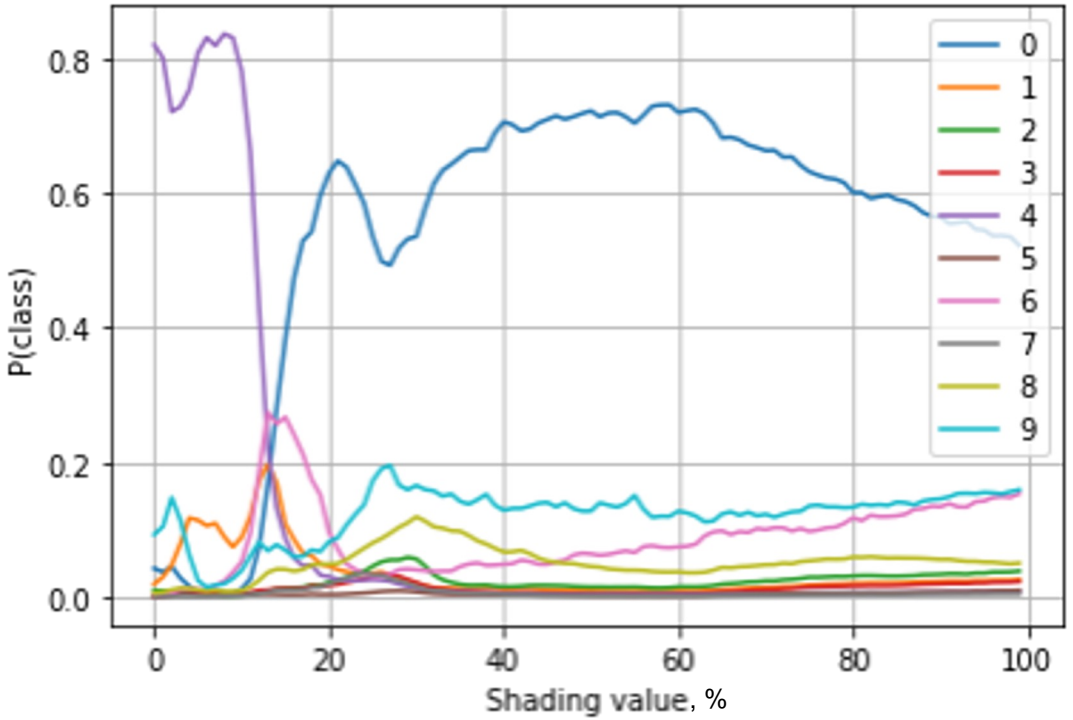
Скачки противоречат гипотезе о том, что SHAP указывает значимые области.

Последовательное затенение

Пример на алгоритме Grad-Cam



Grad-Cam



Затемнение пикселей по степени значимости, вычисленной с помощью Grad-Cam. Каждая линия это вероятность соответствующего класса.

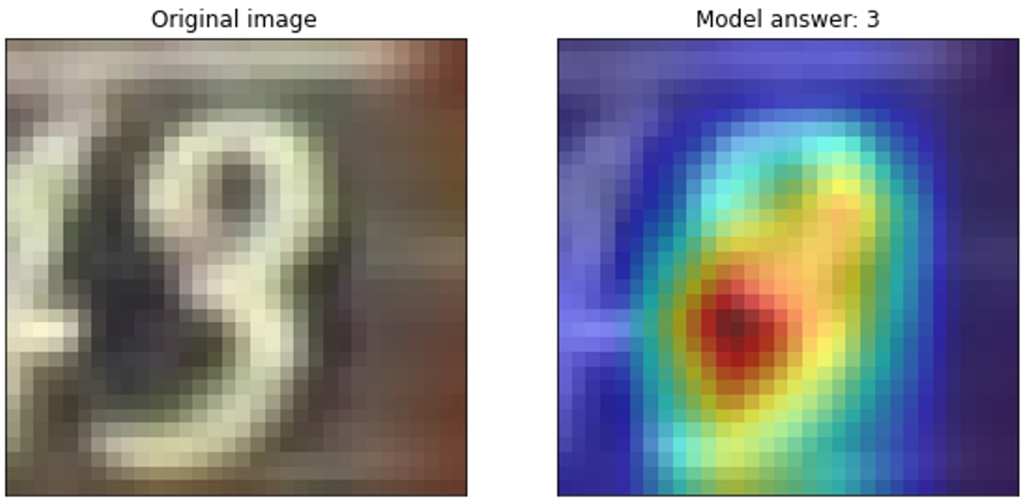
AIRI | Университет Сириус

Попутные выводы

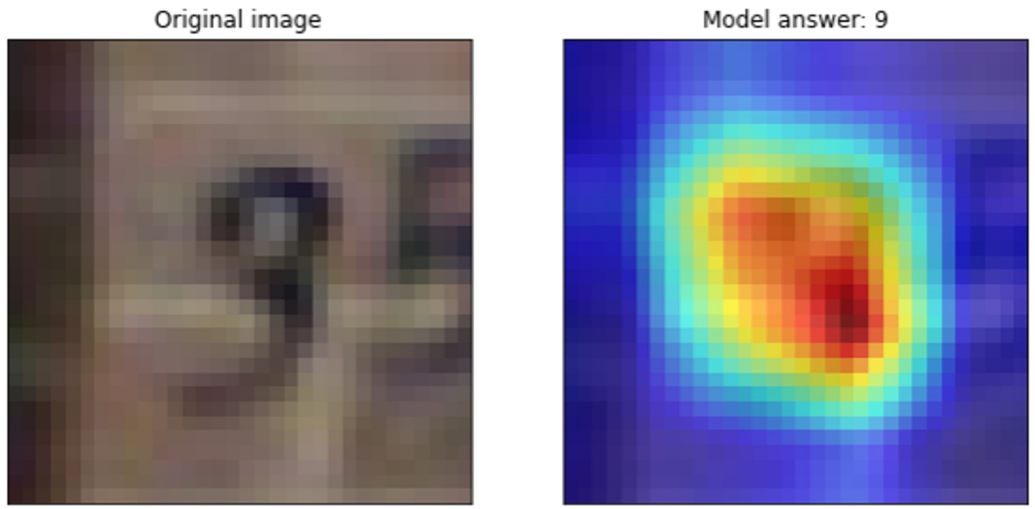
- На интерпретациях SHAP и LIME мы не увидели фокус модели на самих цифрах. Также данные методы не прошли “простые” методы проверки.
- Grad-Cam визуально дает информативные heatmap.
- Grad-Cam также более логично ведет себя по графикам постепенного затемнения.

→ поэтому далее пользуемся именно Grad-Cam

Интерпретация Grad-Cam

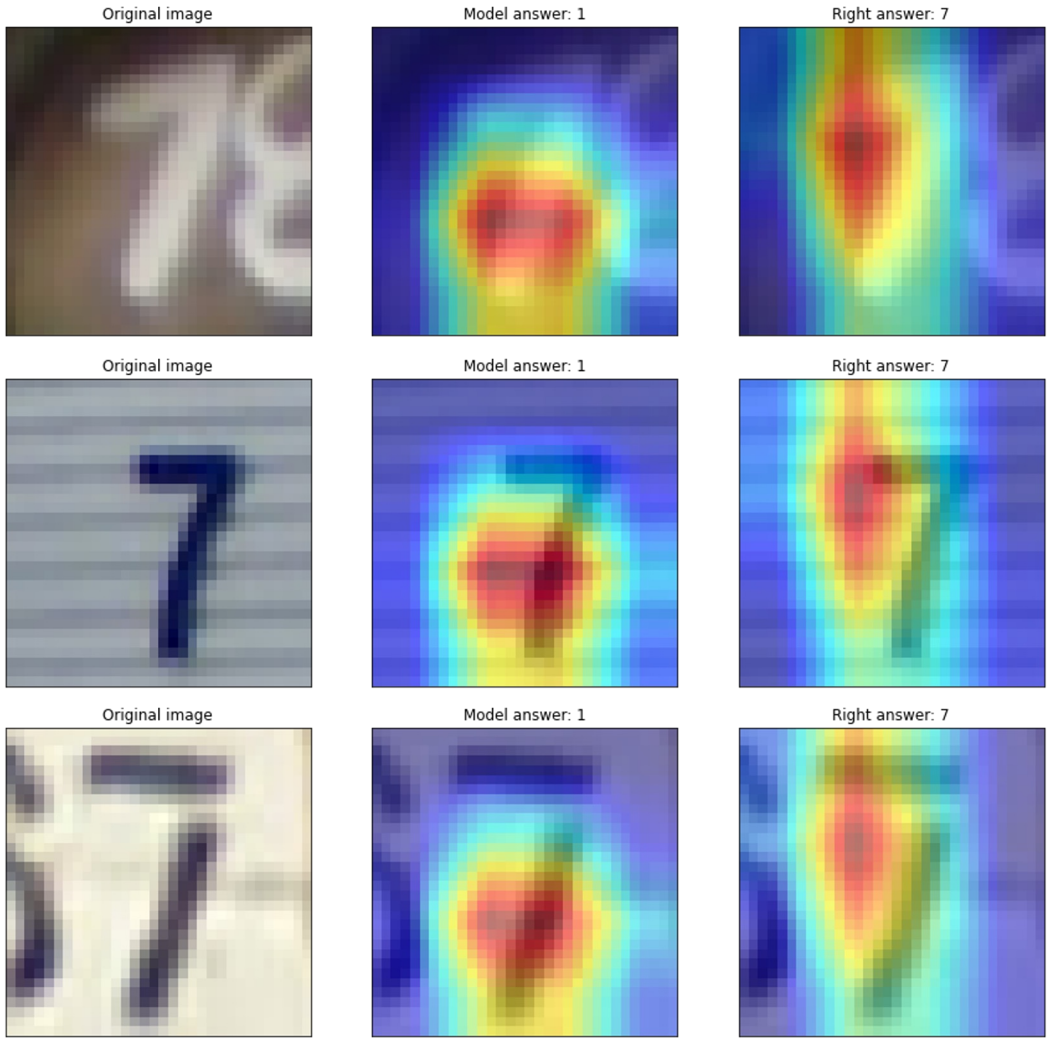


верное предсказание 3

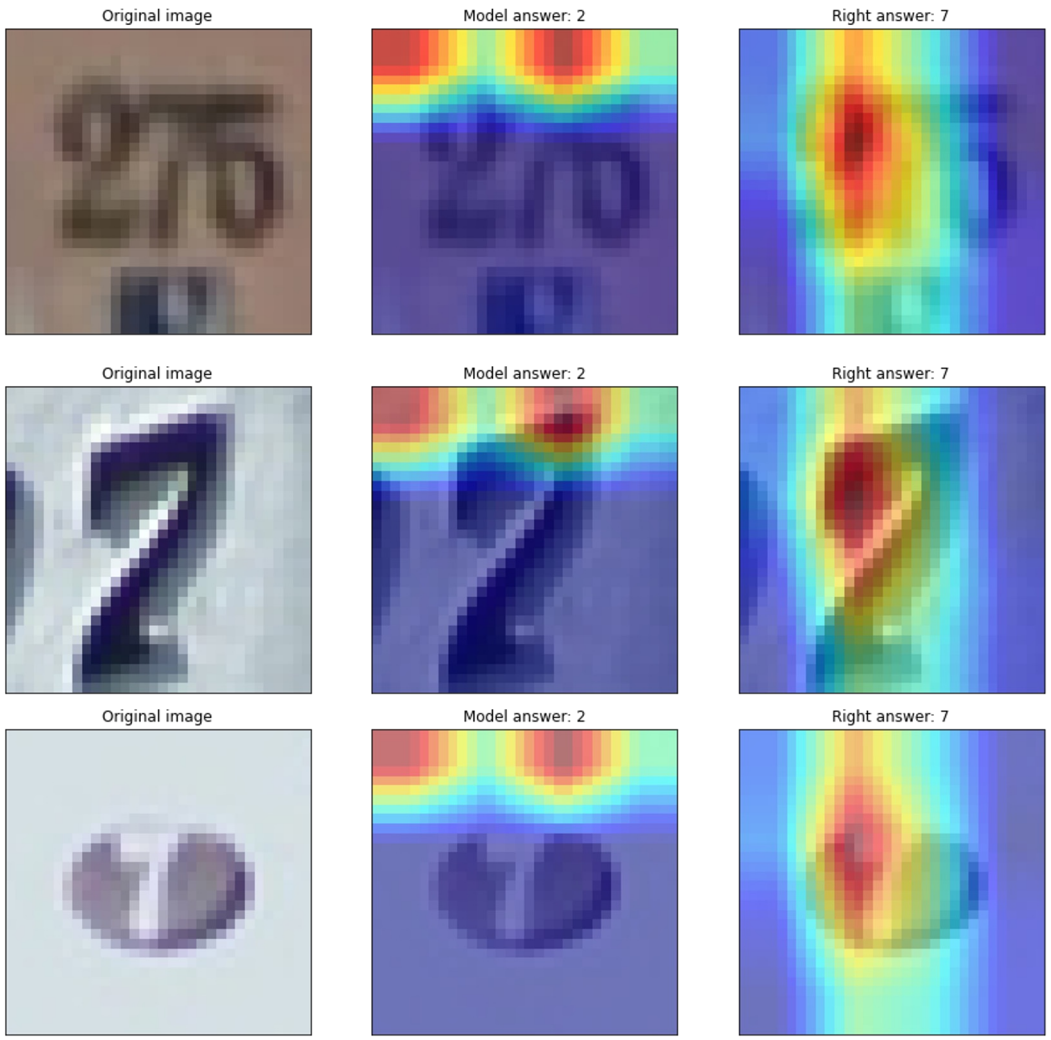


НЕверное предсказание 3

Исправляем предсказание на цифре 7



видит 1, концентрируясь на центральной вертикальной палочке в 7



видит 2, концентрируясь на двух верхних точках картинки

AIRI | Университет Сириус

Самый полезный вывод

В ходе интерпретации результатов обнаружена местами неверная разметка датасета.

Спасибо за внимание!

Забродина Татьяна

Шокоров Вячеслав

Colab code link

AIRI | Университет Сириус