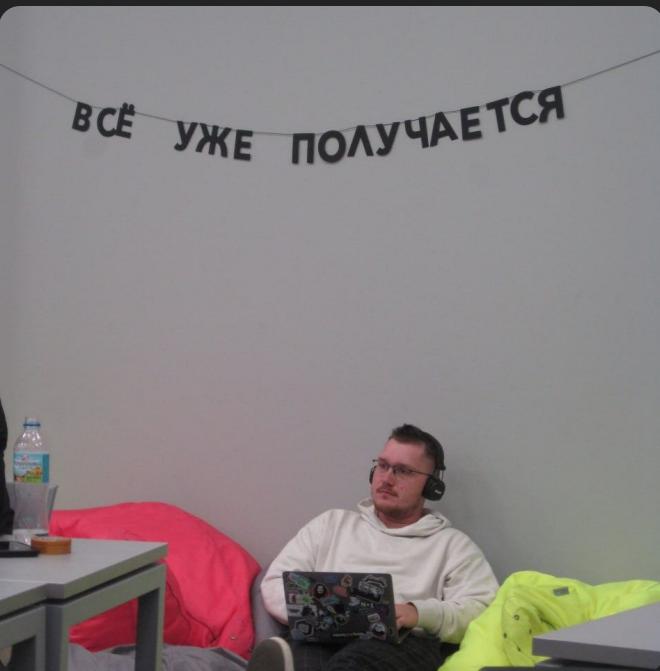


MISIS KAIF

Мастер кадра (МТС)

Команда



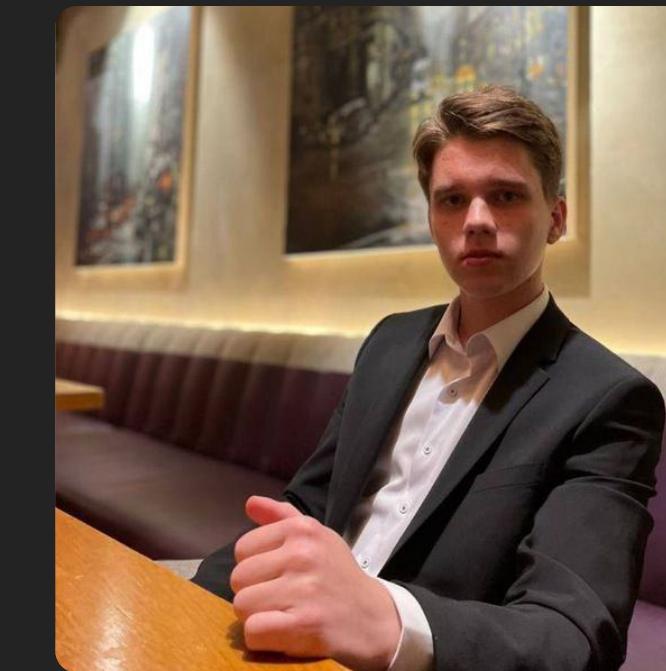
Александр
Горин

ML

LEAD

ML and DS specialist
НИТУ МИСИС
Москва

@ALEKSANDERGORIN



Даниил
Зиновьев

DL

ML at XLABS
НИТУ МИСИС
Москва

@deeeeeenchik



Максим
Хандусь

ML

DATA ANALYST

NLP
НИТУ МИСИС
Москва

@duck_mood

Данные

*парсинг 20 000 фото с открытых
источников*

поиск на сайтах отелей

дедублицирование

поиск готовых датасетов

Разметка

Ансамбль разметчиков

llava:v1.5
llava-llama3
llava-phi3
llava-7b
llava-13b



промпты



voting classifier

в **73%** случаев как минимум 3 модели соглашалась



размеченные данные



эксперименты

```
prompts = [  
    'Exterior',  
    'Room',  
    'Pool',  
    'Billiard',  
    'Bathroom',  
    'Restaurant',  
    'Lobby',  
    'Beach',  
    'Corridors and stairs',  
    'Food',  
    'Conferenceroom',  
    'Gym',  
    'Balcony',  
    'Terrace and patio',  
    'Spa and sauna']
```

+ Code + Markdown
6]: prompts.append(' not '.join(prompts))

Максим

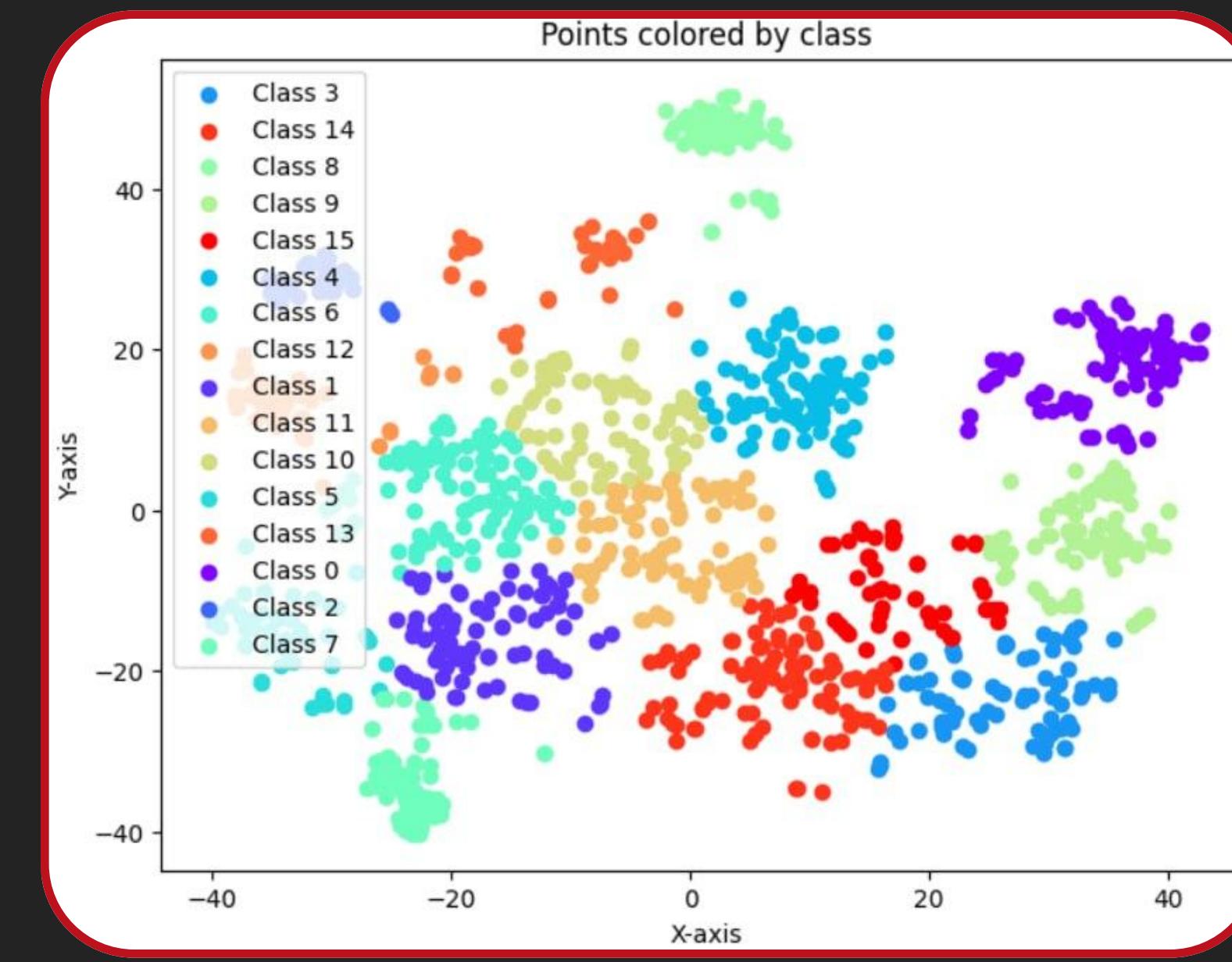
USER: <image>
Classify this image into exactly one of the following 15 categories based on the content of the image:

- Exterior (e.g., building facade, outdoor environment),
- Room (e.g., bedroom, living room, general indoor space),
- Pool (e.g., swimming pool area),
- Billiard (e.g., pool table, billiards room),
- Bathroom (e.g., shower, sink, toilet),
- Restaurant (e.g., dining tables, food service area),
- Lobby (e.g., hotel or office lobby, reception area),
- Beach (e.g., sand, ocean, beach chairs),
- Corridors and stairs (e.g., hallways, staircases),
- Food (e.g., plated meals, buffet),
- Conference room (e.g., meeting room, office chairs, tables),
- Gym (e.g., fitness equipment, workout area),
- Balcony (e.g., small outdoor space attached to a room),
- Terrace and patio (e.g., outdoor seating area, rooftop),
- Spa and sauna (e.g., hot tubs, sauna rooms).

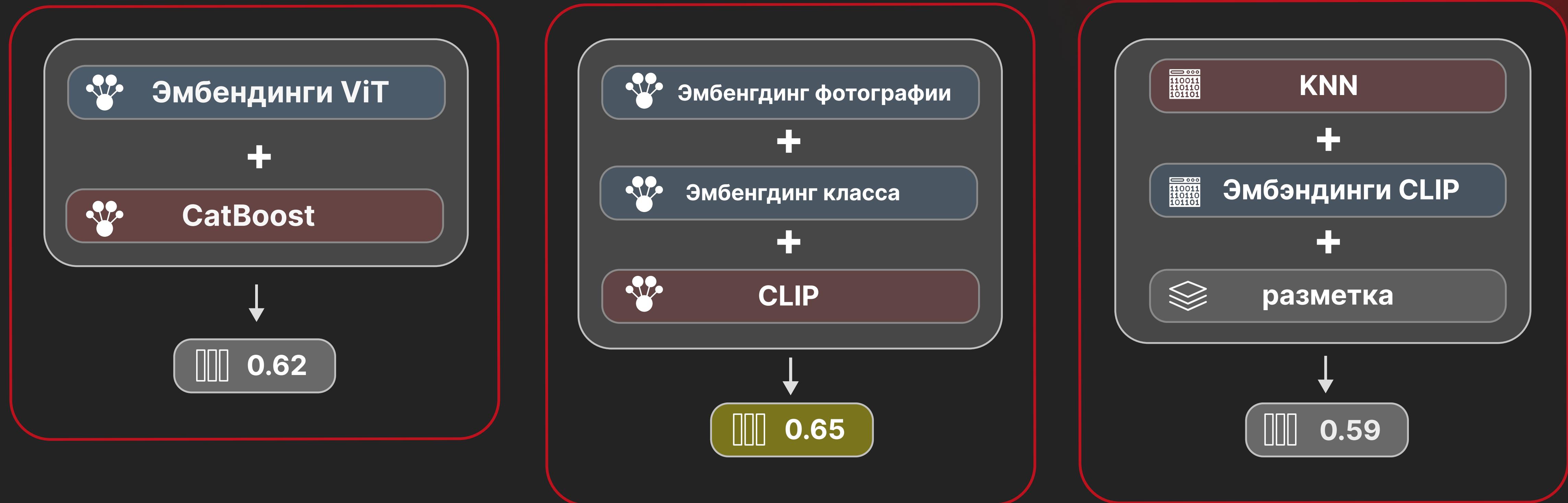
Select the category that best fits the overall scene in the image.
ASSISTANT:

00:02

EDA



Бейзлайны



Обучение

Resnet



Полное обучение



III 0.78

ViT



Полное обучение



III 0.79

CLIP + Classify



CLIP от FLUX



III Последний блок

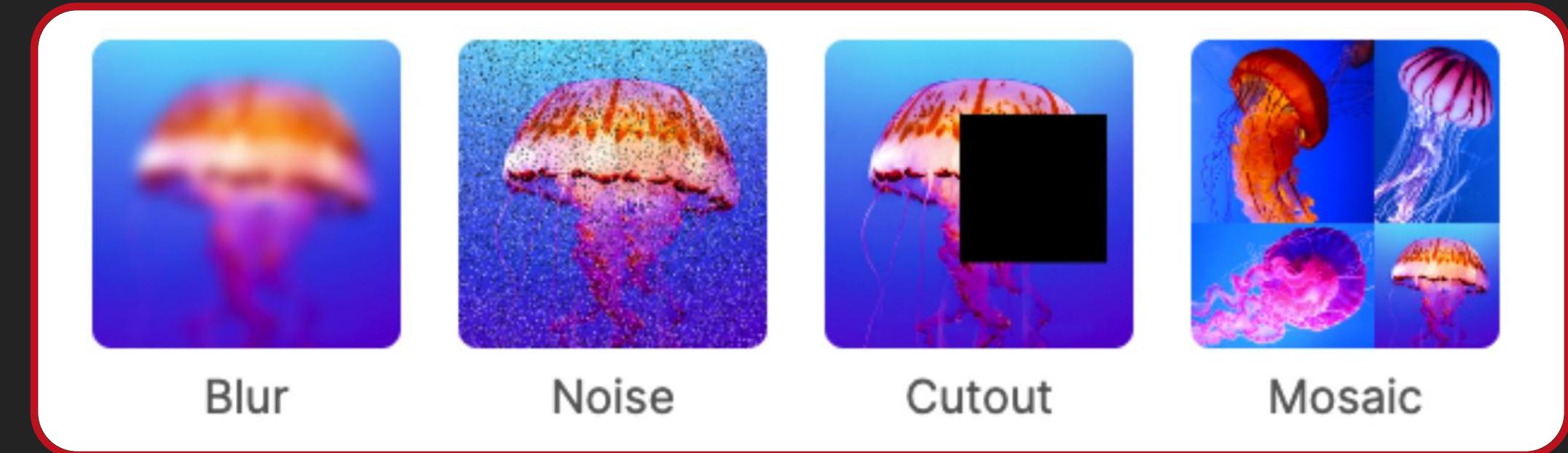
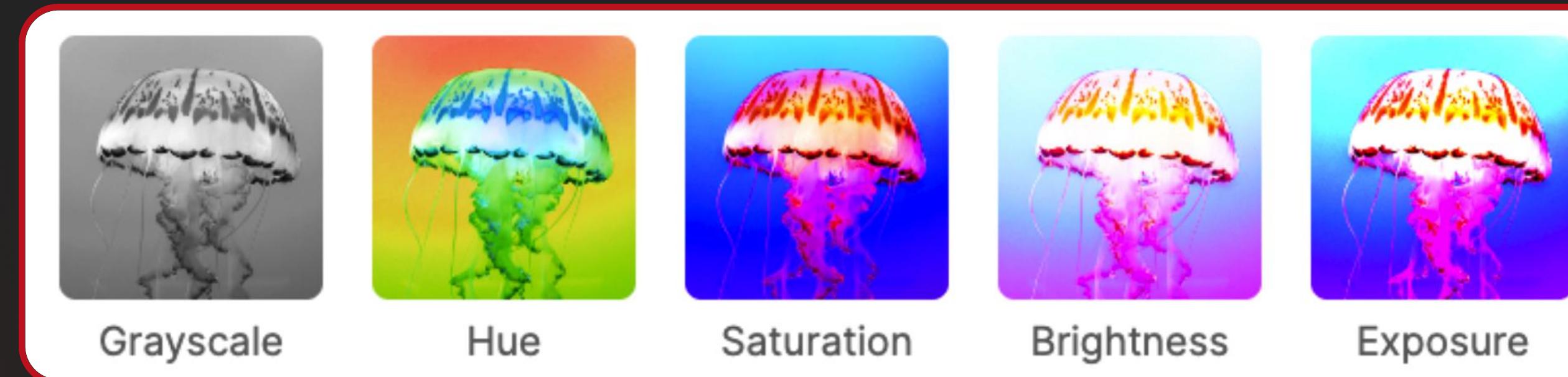
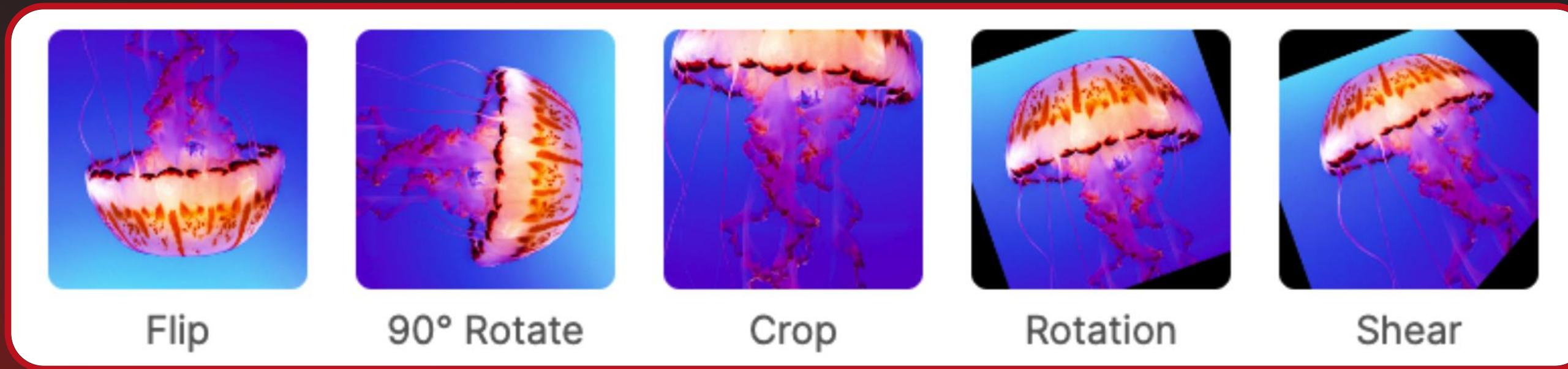


Классификационная
голова



□ 0.82

Аугментации



Дальнейшее развитие

Не успели:

- последняя моделька *image2text* Молмо
- ансамбль предиктов 3 лучших моделей (*resnet, vit ...*)

Не завелось:

- синтетика *image2image* (скор был плохой)
- балансировка классов (стратификация делала хуже), исходное было лучше

Спасибо за внимание!

Наша команда с радостью ответит на ваши вопросы