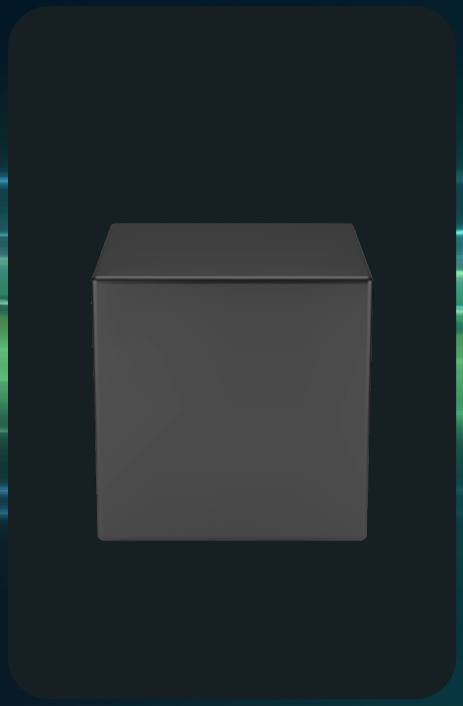
Video Highlights Detection

Александр Гордеев
Computer Vision Engineer
SberDevices

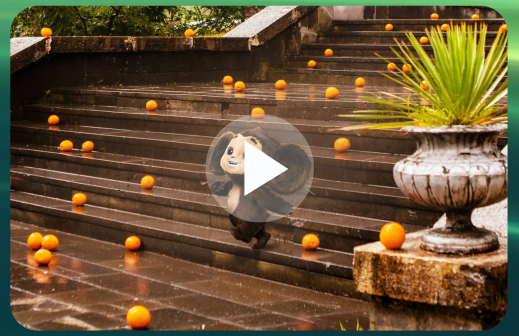
Описание задачи

1 Полнометражный фильм

2 Black box



3 Интересные моменты



Video shorts
здесь и сейчас

SBER DEVICES

Модальности фильма

Аудио модальность

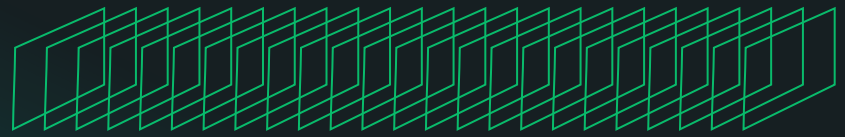
Видео модальность

SBER DEVICES

Структура видео-потока

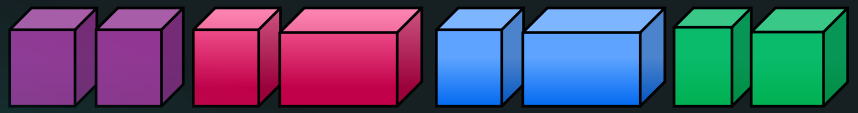
Frame

Одиночное неподвижное изображение, которое при последовательном воспроизведении с другими кадрами видеоролика создает движение на поверхности воспроизведения



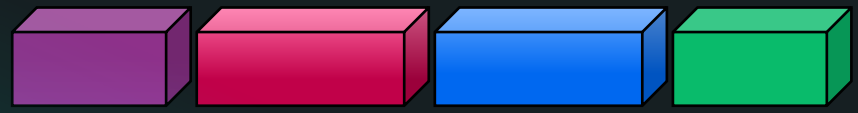
Shot

Серия кадров, снятых одной и той же камерой в течение непрерывного периода времени



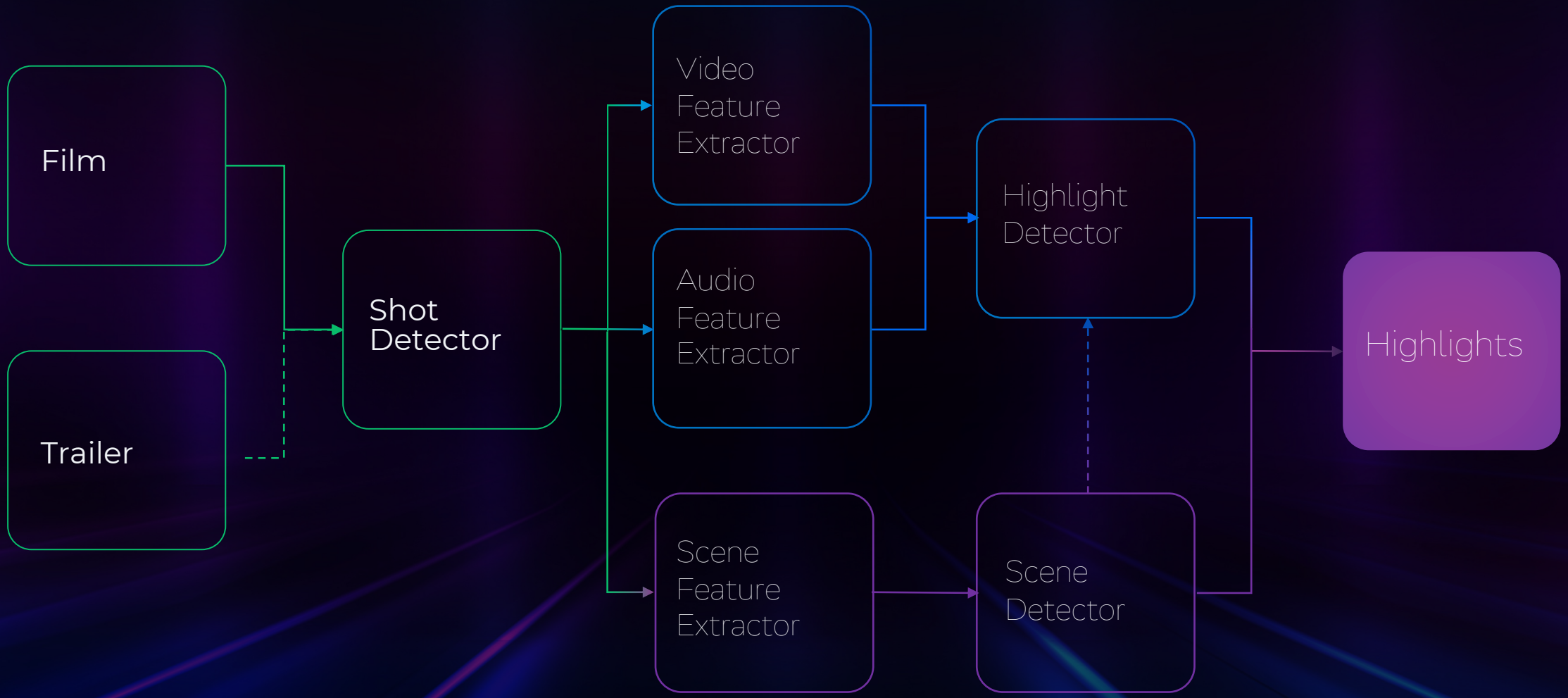
Scene

Серия shot-ов, представляющая собой семантически связную часть сюжета



SBER DEVICES

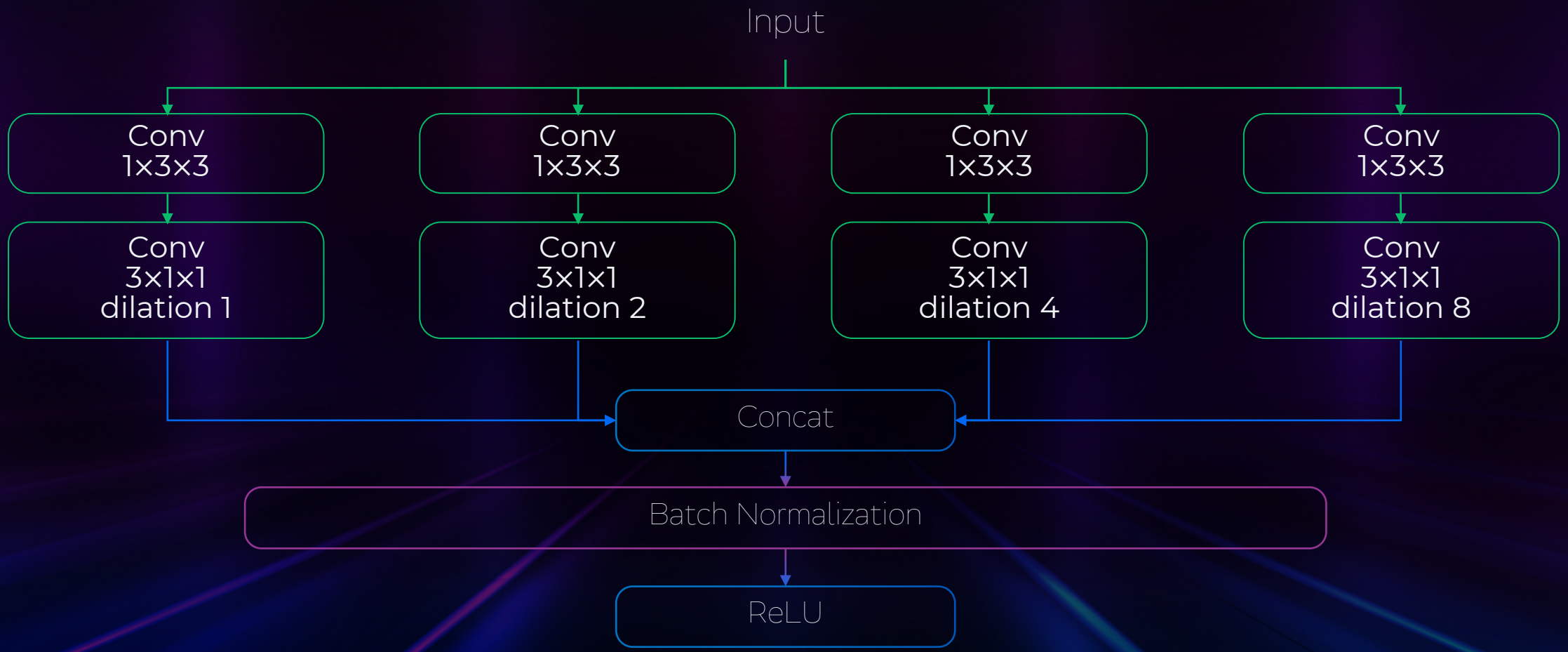
Архитектура решения



SBER DEVICES

Shot Detector

Dilated 3D Conv Block



SBER DEVICES

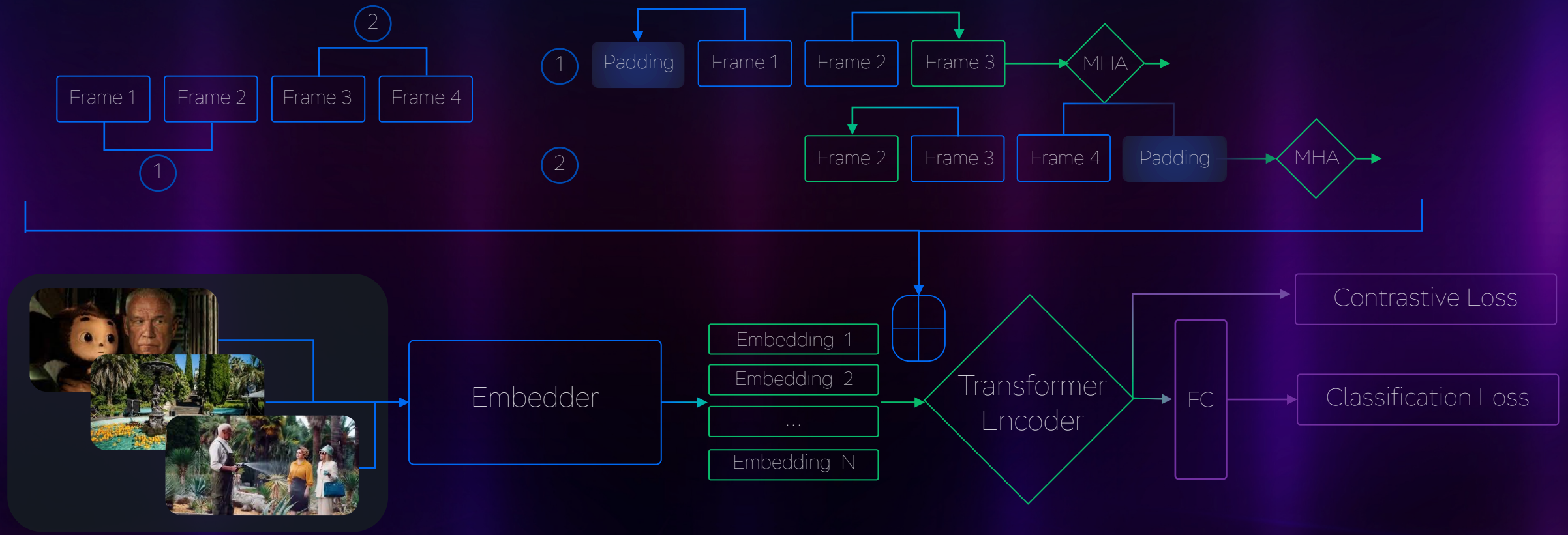
Shot Detector

Бенчмарк:
целевая метрика –
F1-Score

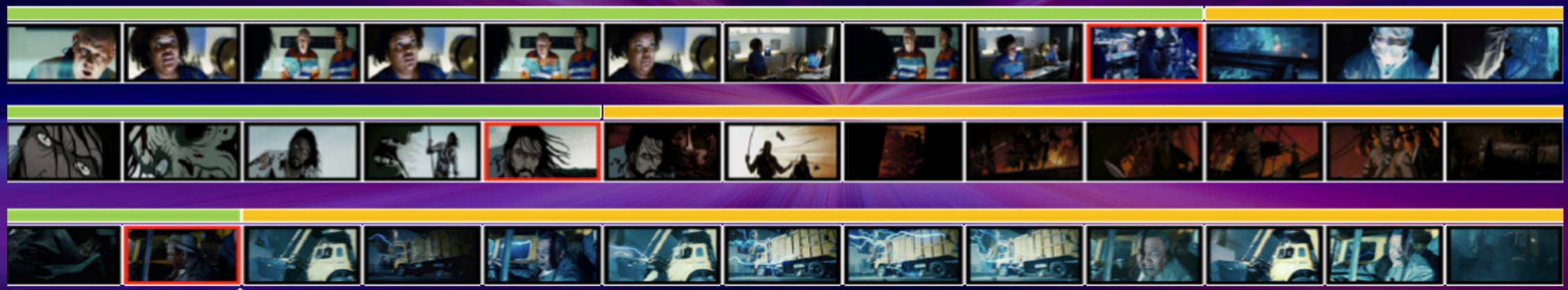
Dataset	ClipShots	BBC	RAI
DSMs	0.761	0.893	0.928
ST ConvNets	0.759	0.926	0.939
TransNet	0.735	0.929	0.943
TransNetV2	0.776	0.962	0.939
AutoShot	0.787	0.971	0.955
Ours	0.808	0.980	-

SBER DEVICES

Shot Detector



Scene Detector



Pseudo-boundary

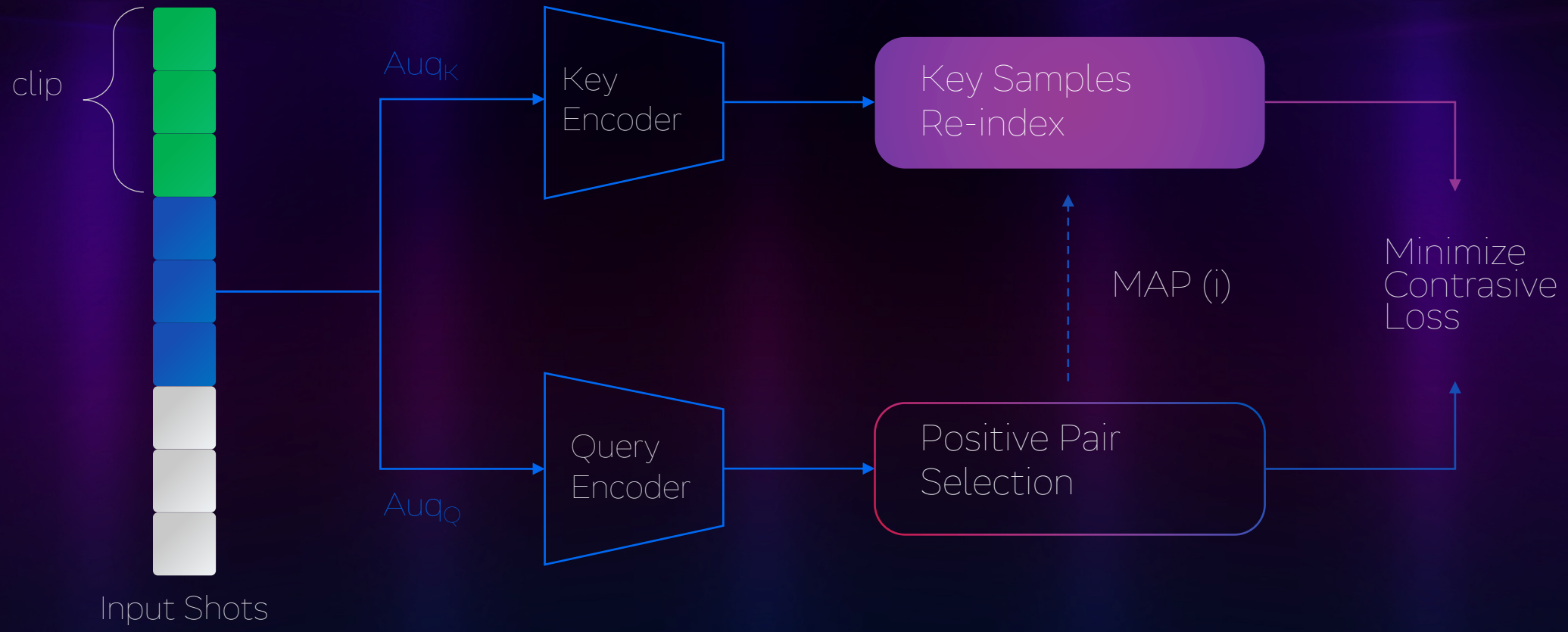
Pseudo Scene A

Pseudo Scene B

SBER DEVICES

Scene Detector

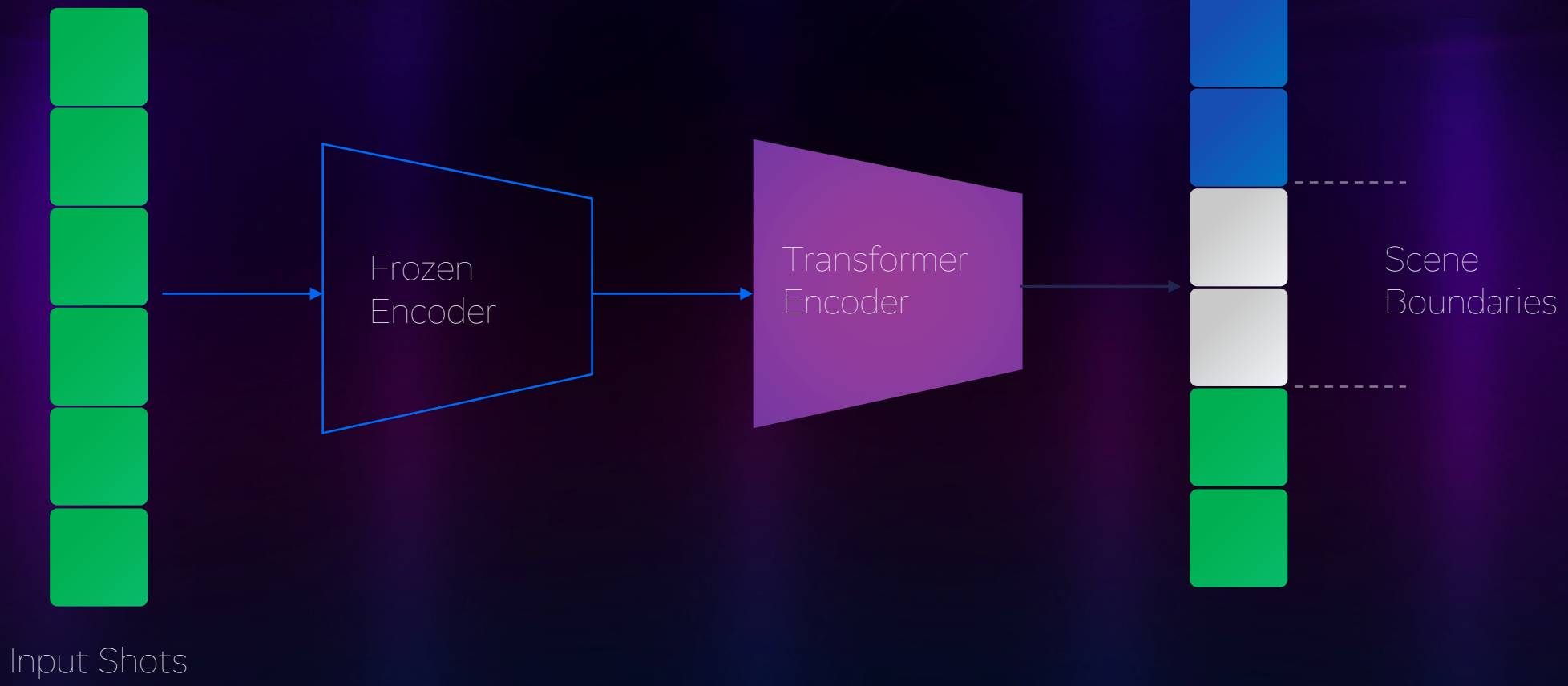
Representation Learning Stage



SBER DEVICES

Scene Detector

Video Scene Segmentation Stage



SBER DEVICES

Highlight Detector

Video

$\{v_i\}_{i=1}^{N_v}$

Video Encoder

Uni-modal Encoder

Audio

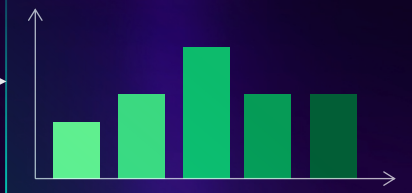
$\{a_i\}_{i=1}^{N_v}$

Audio Encoder

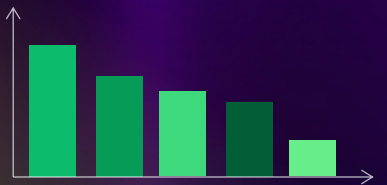
Uni-modal Encoder

Cross-modal Encoder

Contrastive Attention



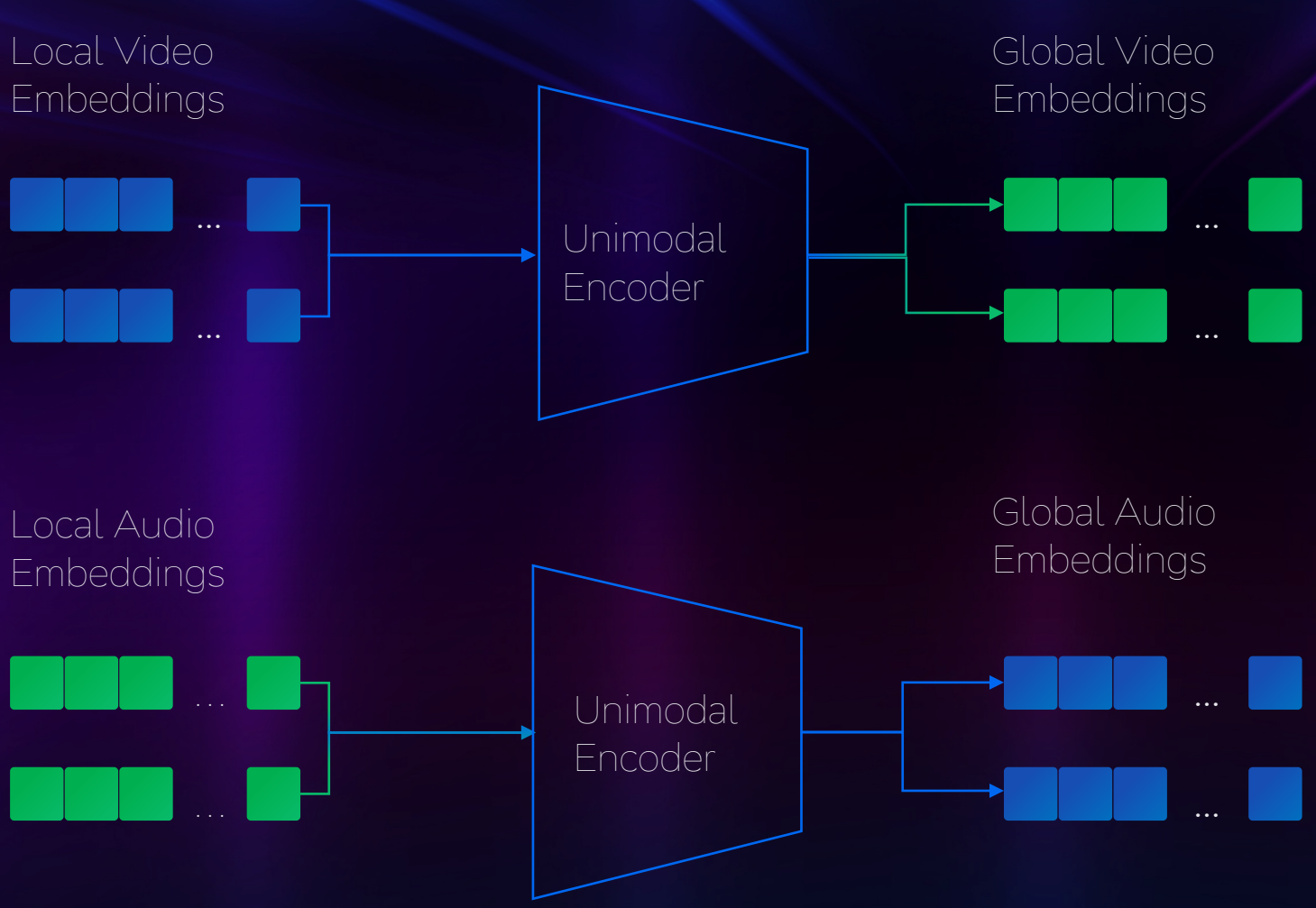
Classification Loss

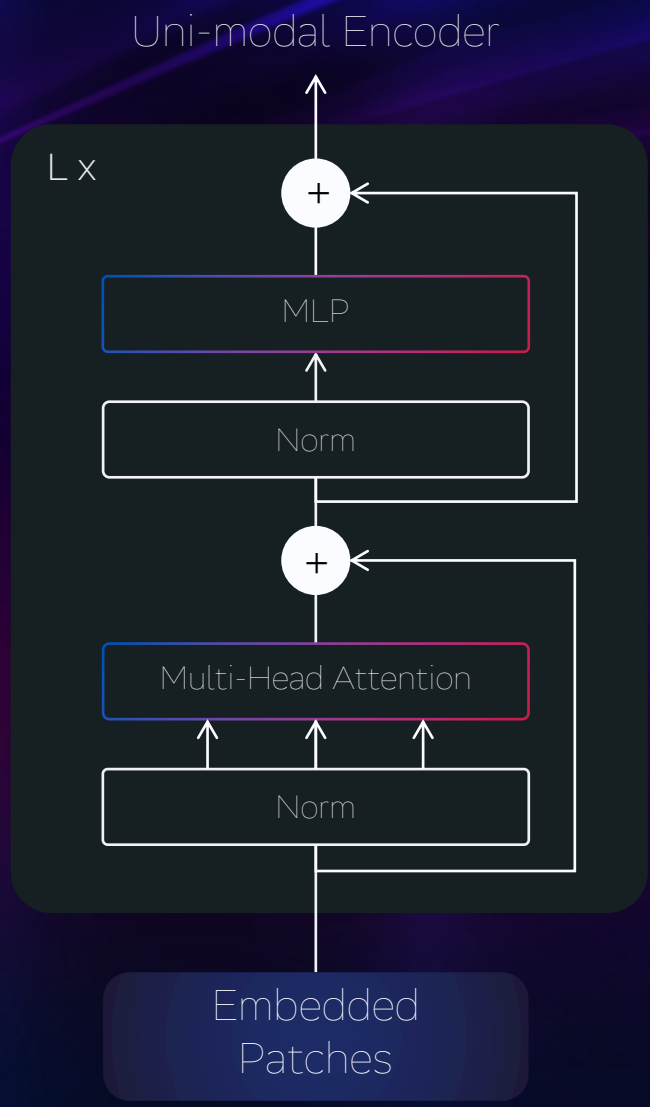


Ranking Loss

SBER DEVICES

Глобальный контекст

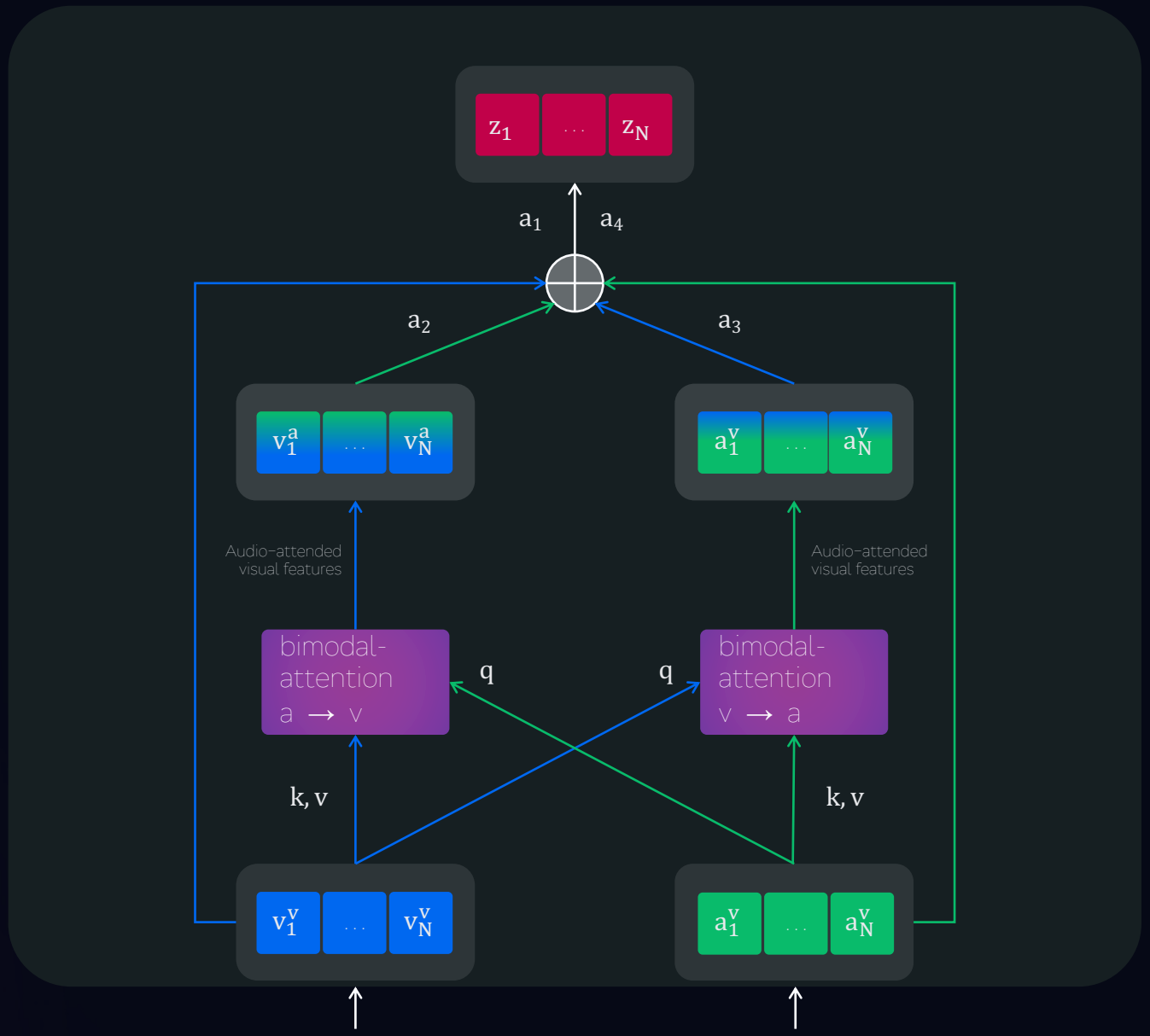




SBER DEVICES

Фьюзинг модальностей

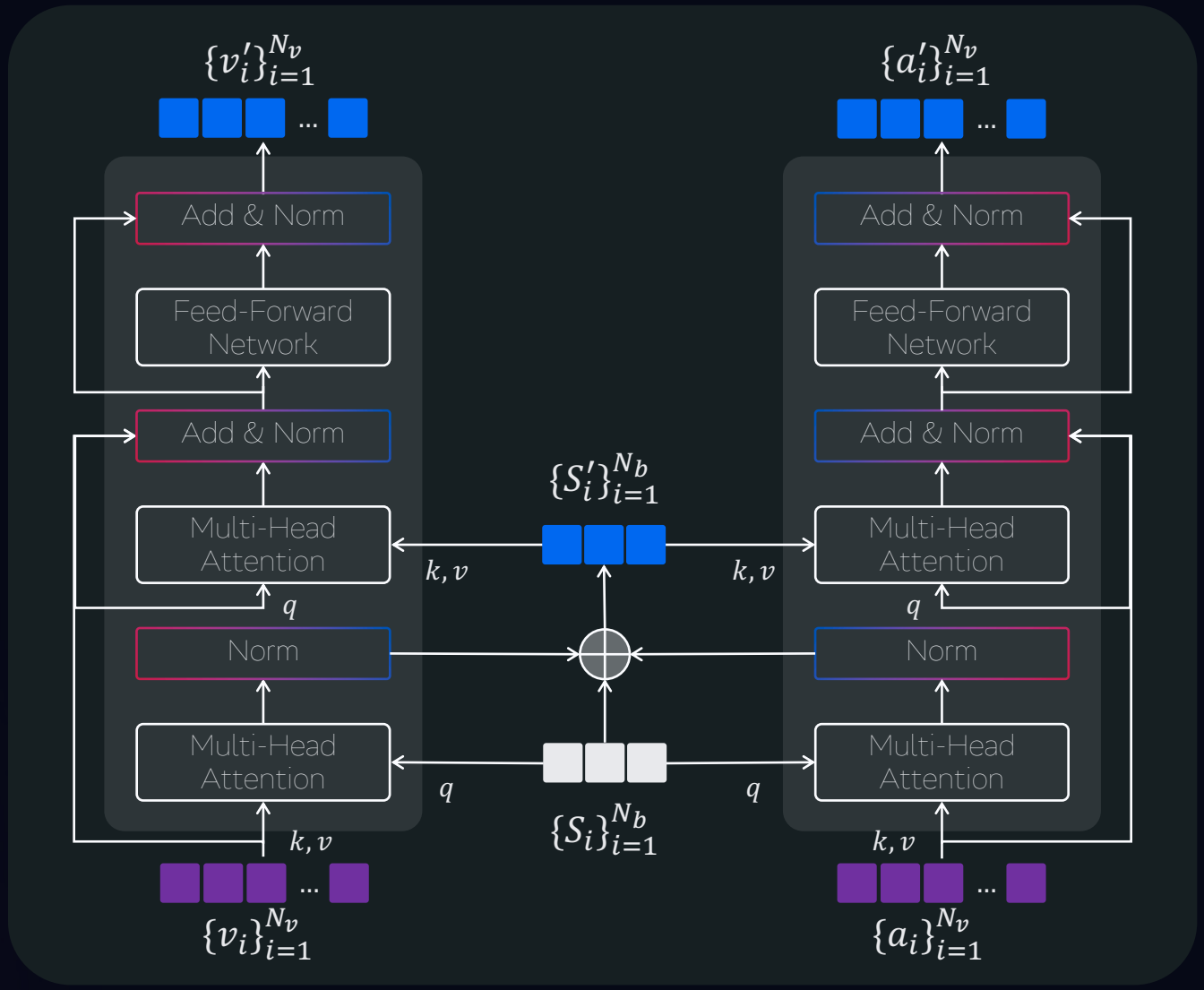
Cross-modal attention



SBER DEVICES

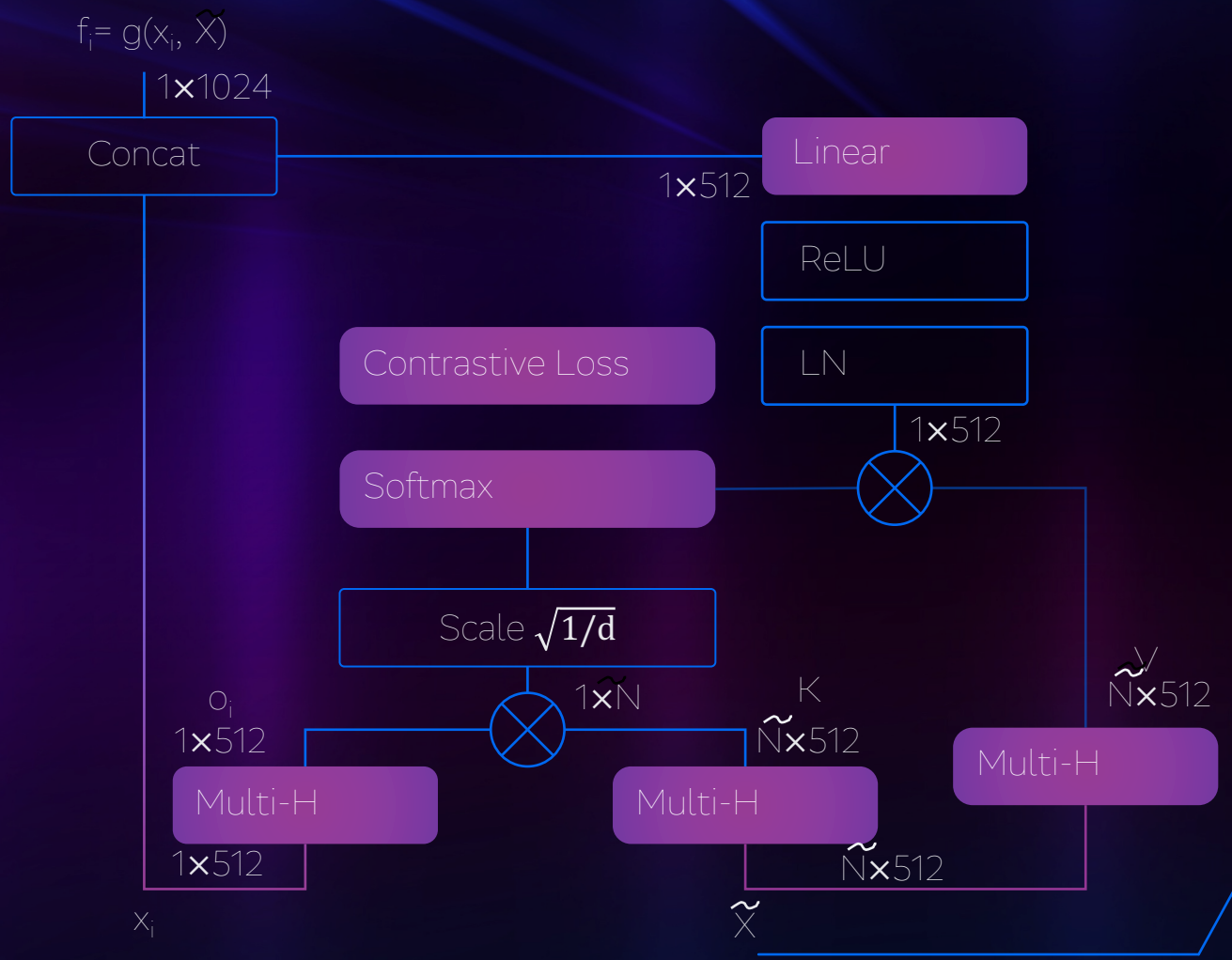
Фьюзинг модальностей

Bottleneck transformer module

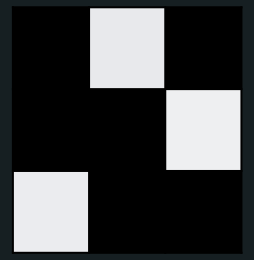


SBER DEVICES

Contrastive attention



Co-Attention block



$$Att_i = \max_{\tau \in N_t} (q_i \circledast m_\tau)$$

SBER DEVICES

Postprocessing

2

2

15

5

23

41

7

12

32

8

5


SBER DEVICES