

# EASY AI —— 具体业务中的 端侧 AI 创新

张静媛

百度资深研发工程师



# 大纲

---

- Paddle.js 2.0
- Web 视频会议背景替换
- 开放平台
- 未来展望

# Web AI 发展趋势 & 优势

## ➤ 发展趋势

1. 算力增强
2. 模型优化
3. Web 能力扩大

## ➤ 优势

- ✓ 低延迟
- ✓ 低成本
- ✓ 隐私保护

### 算力增强

✓ CPU/GPU通用计

### 算力稳步增强

### 模型优化

✓ 轻量模型

✓ 模型量化

### Web 能力扩大

✓ WebRTC

✓ WebGL

✓ WebGPU

✓ WebAssembly

# Paddle.js 落地场景

➤运营创新



双旦猜丁壳

➤商业类

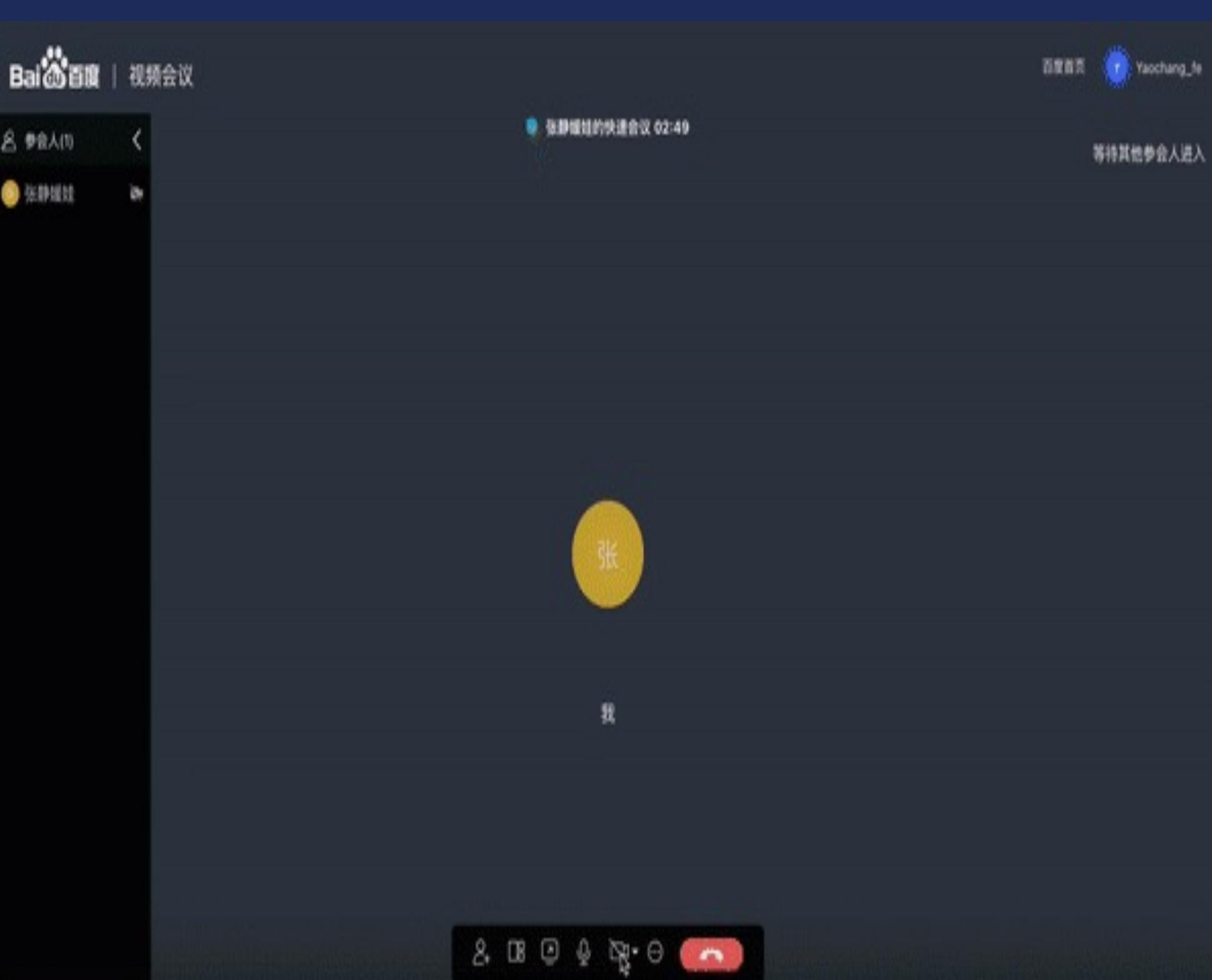


六一



美妆 (探索)

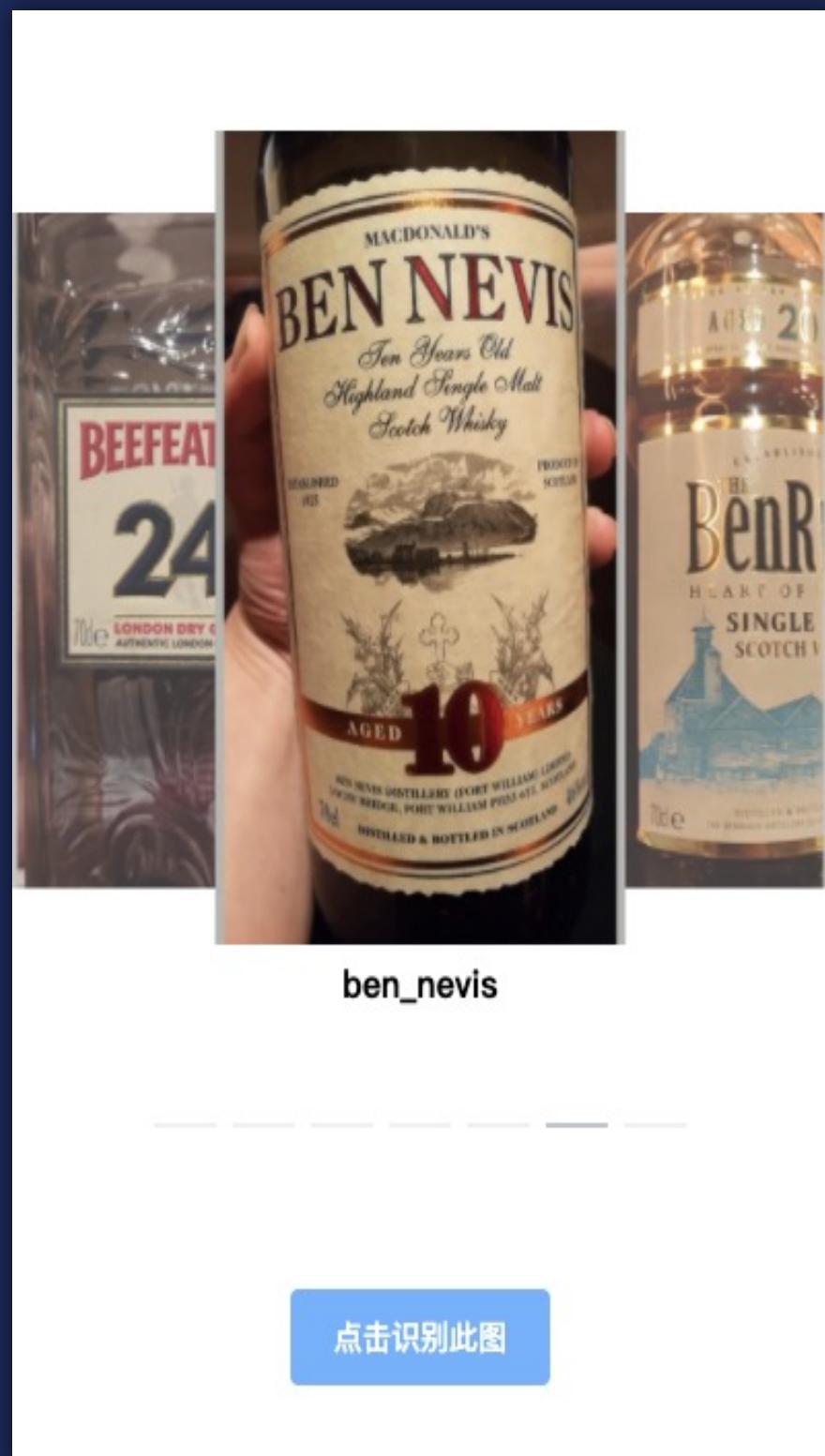
➤智能办公



WEB 视频会议背景替换

# Paddle.js 落地场景

## ➤第三方能力输出



江小白



小视科技（待上线）

## ➤工具类



文本检测

## ➤学习案例



预测结果:

黄反暴恐识别

# Paddle.js 一个高性能、轻量级、灵活性强且易于扩展的前端深度学习推理框架

图像分类

人像分割

OCR

面部识别

目标检测

肢体预测

PaddlePaddle

TensorFlow

ONNX

Caffe

Paddle.js

浏览器

小程序

Node.js

CPU

GPU

# Paddle.js 2.0 全新升级

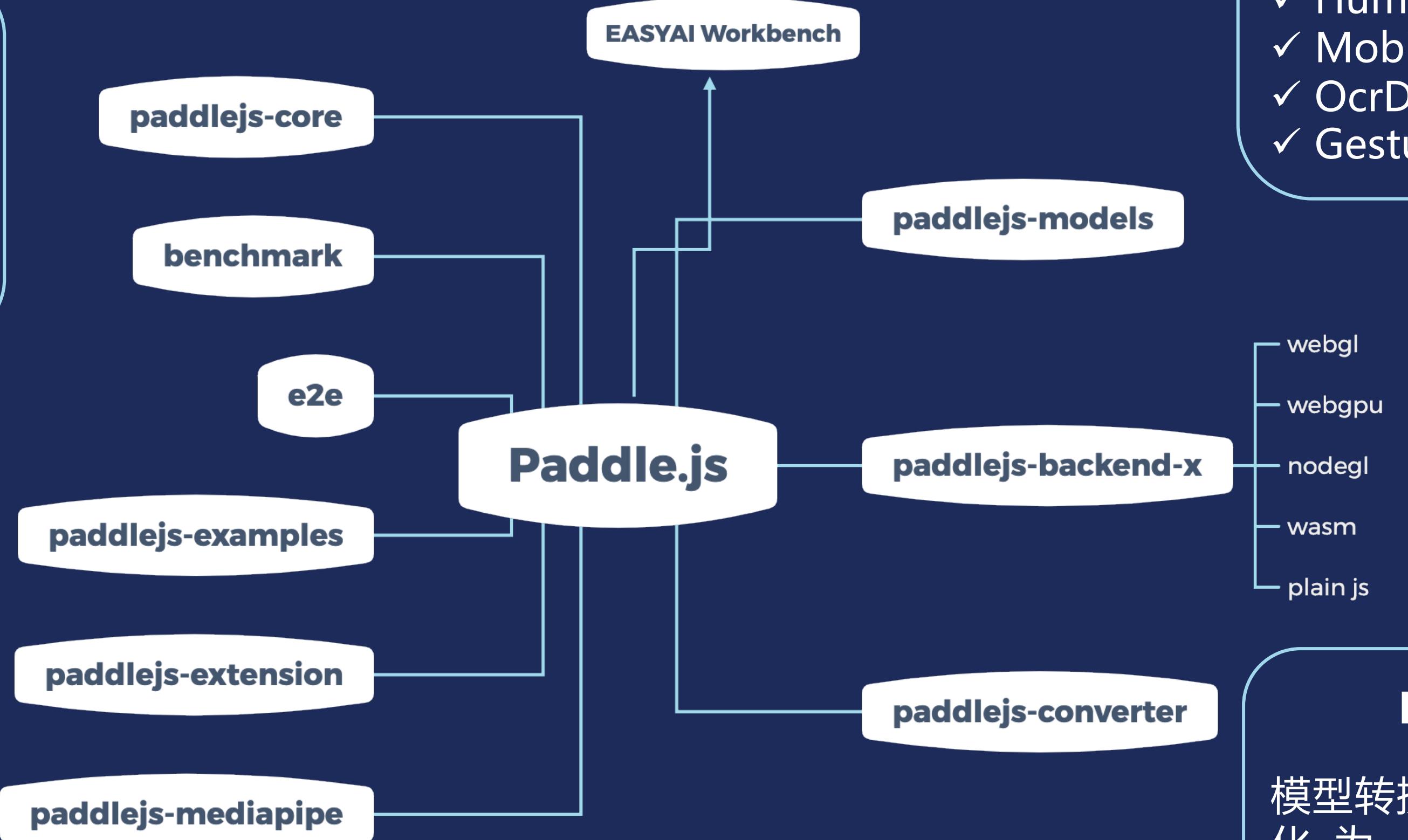
## paddlejs-core

推理引擎的核心部分，负责整个引擎的推理流程运行，提供计算方案注册、环境变量注册

## paddlejs-mediapipe

数据流出工具库

- ✓ Camera
- ✓ Opencv
- ✓ DataProcessor



## paddlejs-models

封装好的模型 SDK，提供开箱即用的 AI 能力

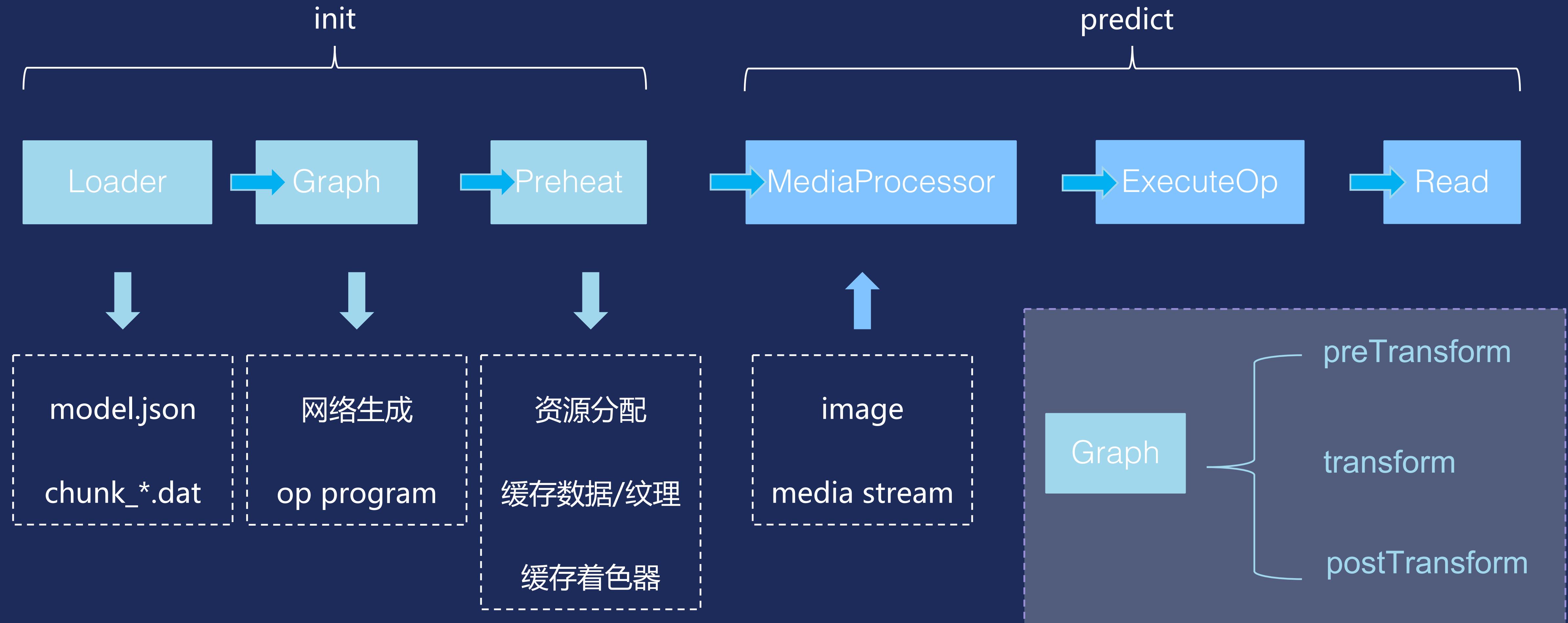
- ✓ Humanseg 人像分割
- ✓ Mobilenet 物品分类
- ✓ OcrDetection 文本检测
- ✓ Gesture 手势识别

## paddlejs-converter

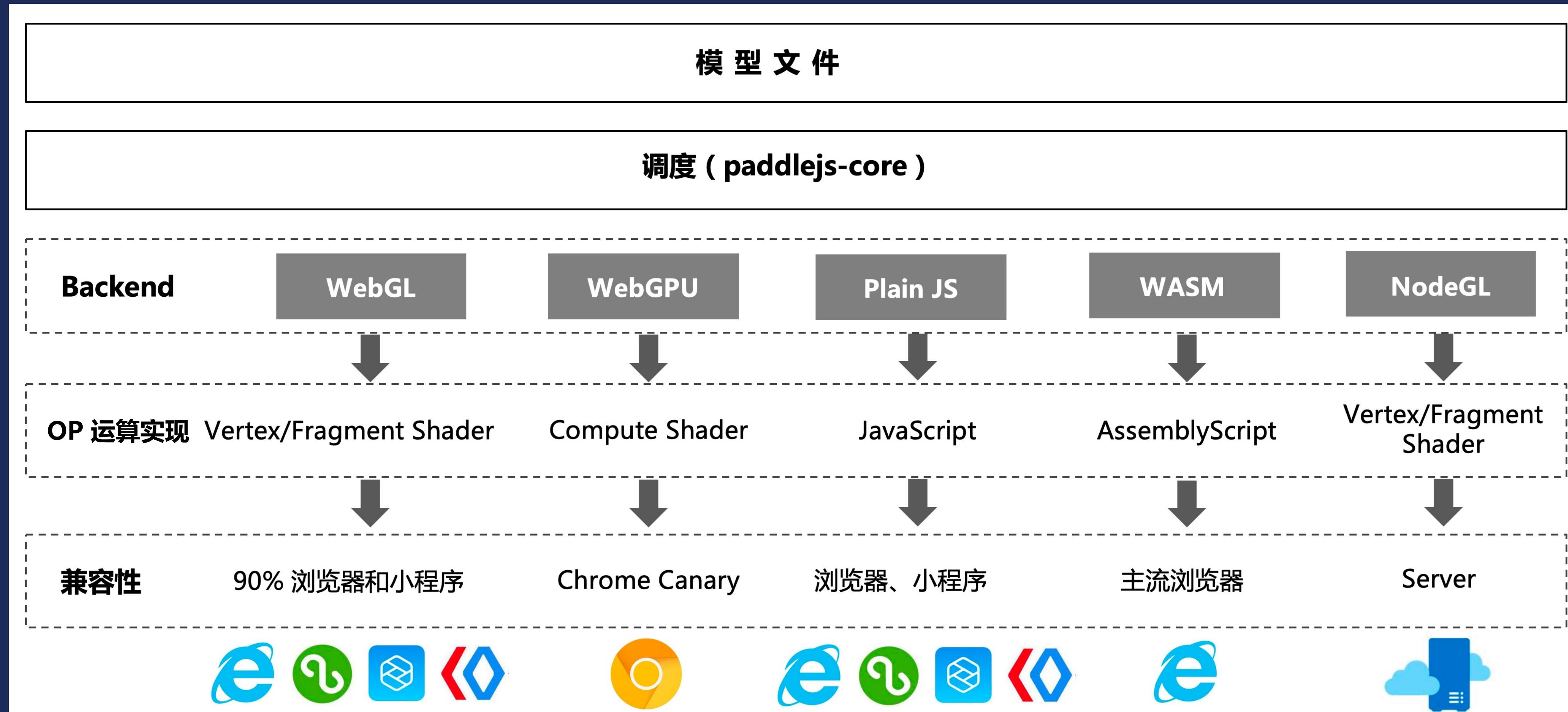
模型转换工具，将 paddle 模型转化为 Paddle.js 模型文件 model.json chunk.dat；支持模型量化和模型融合

webgl  
webgpu  
nodegl  
wasm  
plain js

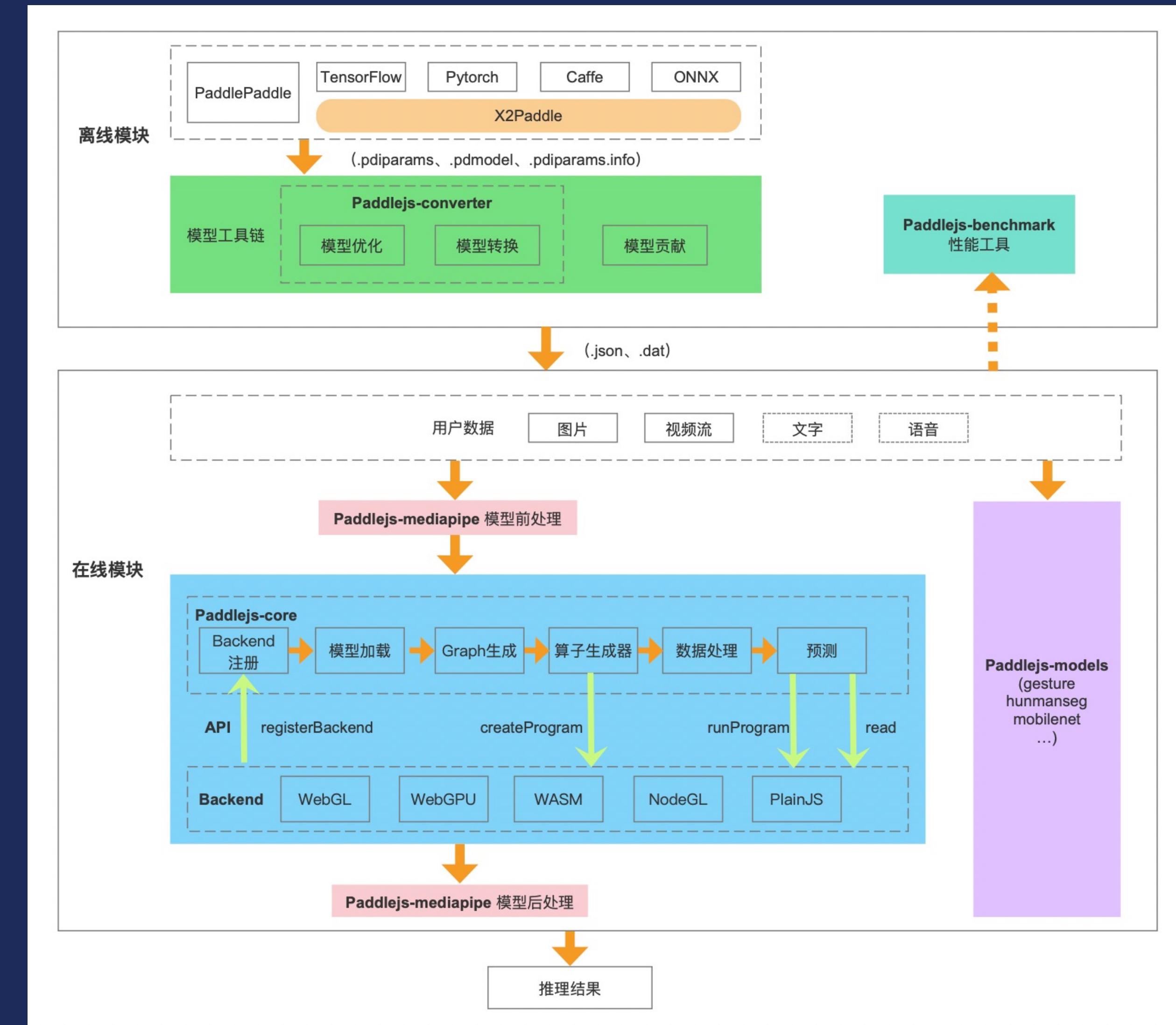
# paddlejs-core



# backend



# Paddle.js 2.0 核心模块



# Paddle.js 2.0 核心指标

## 高性能

OP 融合

FP16 量化

WEBGL\_PACK

MediaPipe

## 易用性

封装 model sdk , 提供开箱即用的

AI 能力

提供 paddlejs-mediapipe 数据流

工具 sdk

## 通用性/兼容性

支持主流轻量模型结构

覆盖所有浏览器

# 高性能

视频流场景

图优化

op 融合、FP 16量化

推理优化

WebGL Pack、Half Float

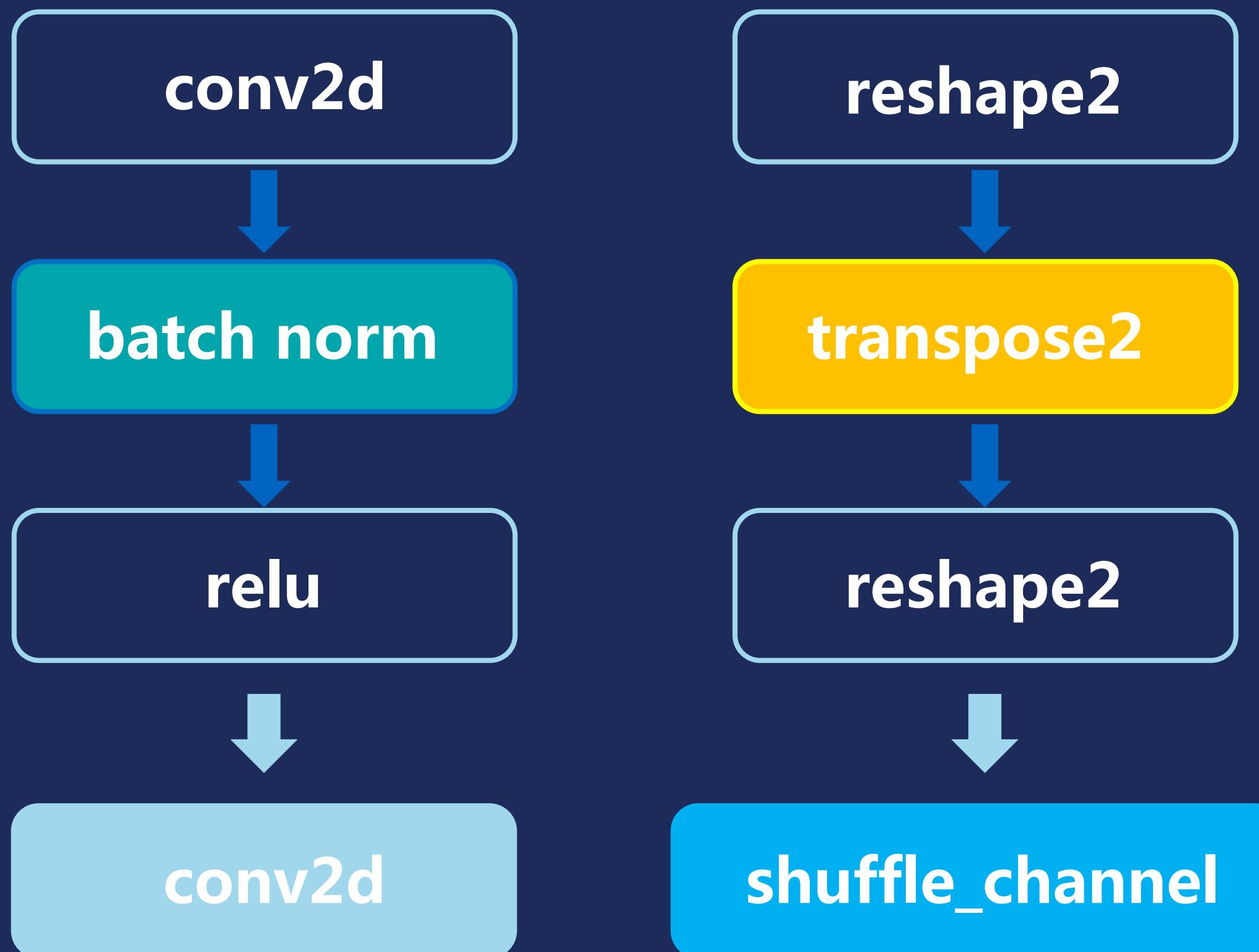
渲染优化

MediaPipe

# 高性能

## OP 融合

减少 I/O、简化图结构



## FP16 量化

FP32 精度模型量化为 FP16 精度模型

减小模型体积，参数文件 -50%

降低 模型加载耗时

加快 模型推理

# 高性能

## GPU 加速

✓ 并行计算 WebGL PACK

✓ CPU迁移 GPU

✓ 半浮点计算

`webgl_pack_channel`

`webgl_feed_process`

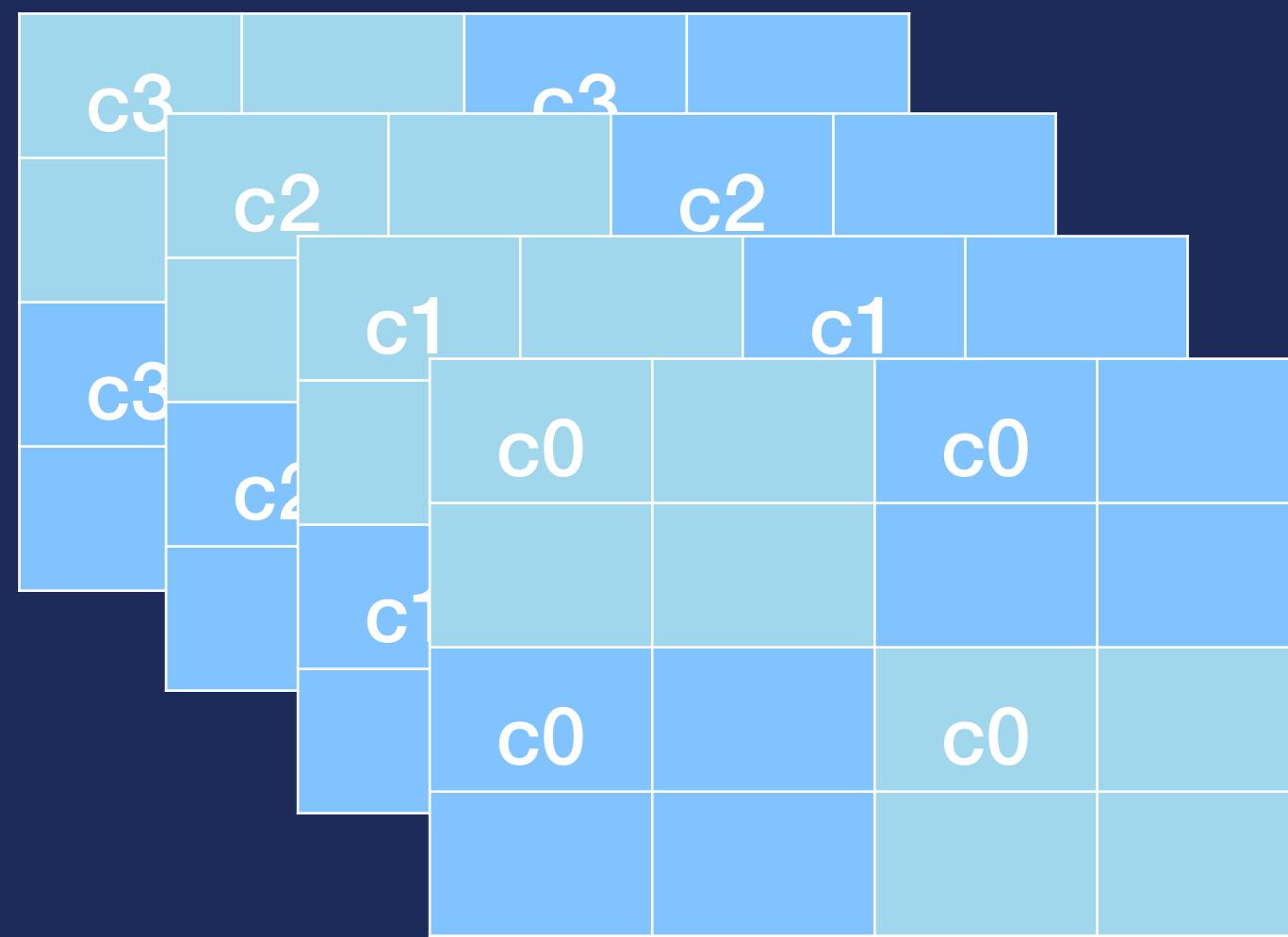
`webgl_force_half_float_texture`

`webgl_pack_output`

# 高性能

webgl\_pack\_channel

mobilenet WebGL 2.0 预测性能提升 39.6%  
humanseg\_mobile WebGL 2.0 性能提升 15%



单通道数据排布

webgl\_pack\_output

humanseg\_398x224 WebGL 2.0 性能提升 9.6%

c0	c1	c0	c1
c2	c3	c2	c3
c0	c1	c0	c1
c2	c3	c2	c3

4 通道数据排布

d0		d1		d2	
d3		d4		d5	

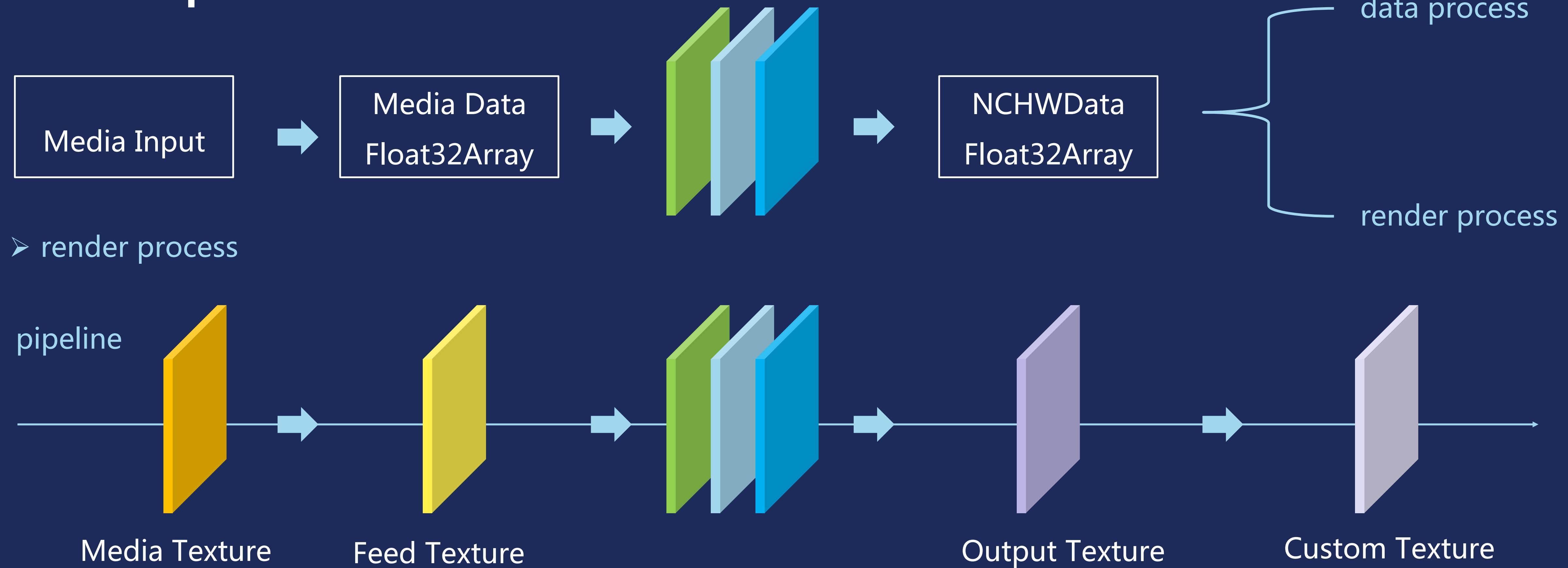
原始输出 [1, 1, 2, 3]

d0	d1	d4	d5		
d2	d3				

pack 输出 [1, 1, 1, 3]

# 高性能

## MediaPipe



# 易用性 —— paddlejs-models

```
import * as humanseg from '@paddlejs-models/humanseg';

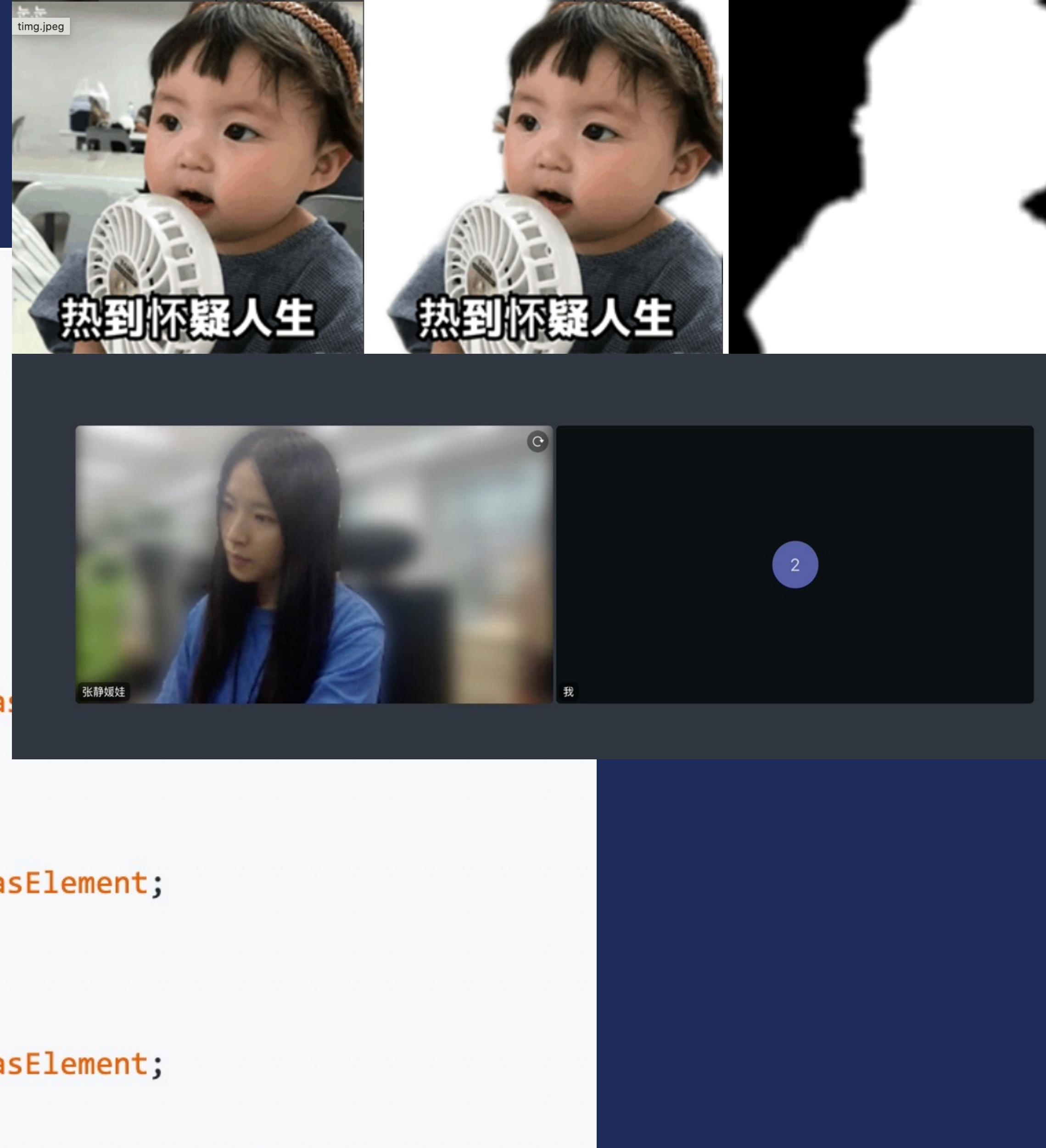
// load humanseg model
await humanseg.load();

// get the gray value [2, 288, 160];
const { data } = await humanseg.getGrayValue(img);

// replace background & draw human segmentation in back canvas
const backCanvas = document.getElementById('back') as HTMLCanvasElement;
humanseg.drawHumanSeg(data, backCanvas);

// blur background
const blurCanvas = document.getElementById('blur') as HTMLCanvasElement;
humanseg.blurBackground(data, blurCanvas);

// draw the background mask
const maskCanvas = document.getElementById('mask') as HTMLCanvasElement;
humanseg.drawMask(data, maskCanvas, true);
```



# 易用性 —— paddlejs-mediapipe

```
import Camera from '@paddlejs-mediapipe/camera';

const option = {
    // video width
    width?: number,
    // video height
    height?: number,
    // mirror or not
    mirror?: boolean,
    // canvas DOM
    targetCanvas?: HTMLCanvasElement,
    // video rendered successfully
    onSuccess?: () => void,
    // video rendering failed
    onError?: NavigatorUserMediaErrorCallback,
    // browser does not support the getusermedia API
    onNotSupported?: () => void,
    // get every frame of video
    onFrame?: (canvas: HTMLCanvasElement) => void
};
```

```
const camera = new Camera(video, option);
// video play
camera.start();
// video pause
camera.pause();
// video stop
camera.stop();
```



# 易用性 —— Env Flag

- webgl\_pack\_channel
- webgl\_pack\_output
- webgl\_feed\_process
- webgl\_force\_half\_float\_texture

```
env.set('webgl_pack_channel', true);  
  
env.set('webgl_pack_output', true);  
  
env.set('webgl_feed_process', true);  
  
env.set('webgl_force_half_float_texture', true);
```

# 通用性 / 兼容性

## 模型结构

主流轻量模型结构

MobileNetV1、MobileNetV2、

## backend

WebGL 1.0 / 2.0、WebGPU

MobileNetV3、MobileNetV3-SSD、

Chrome、Edge、Safari、Firefox

GhostNet、ShuffleNetV2

Plain JS、WASM

百度App、Safari、Chrome、UC ...

HRNet\_w18\_small

NodeGL

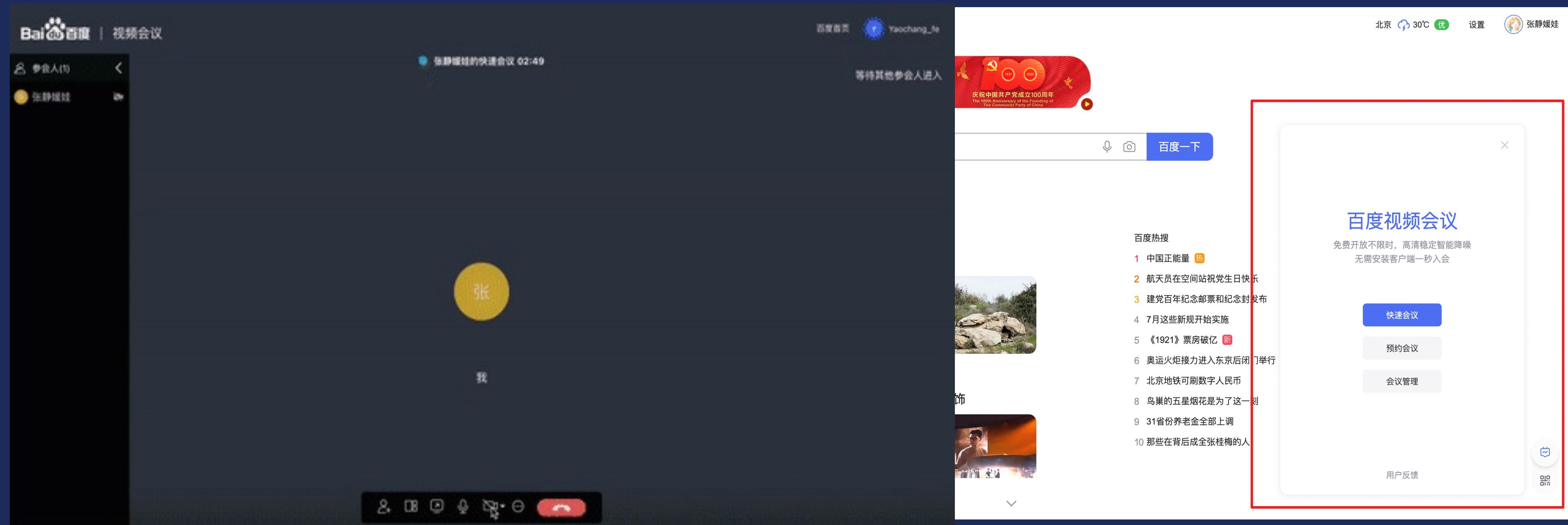
百度智能小程序、微信小程序

# 大纲

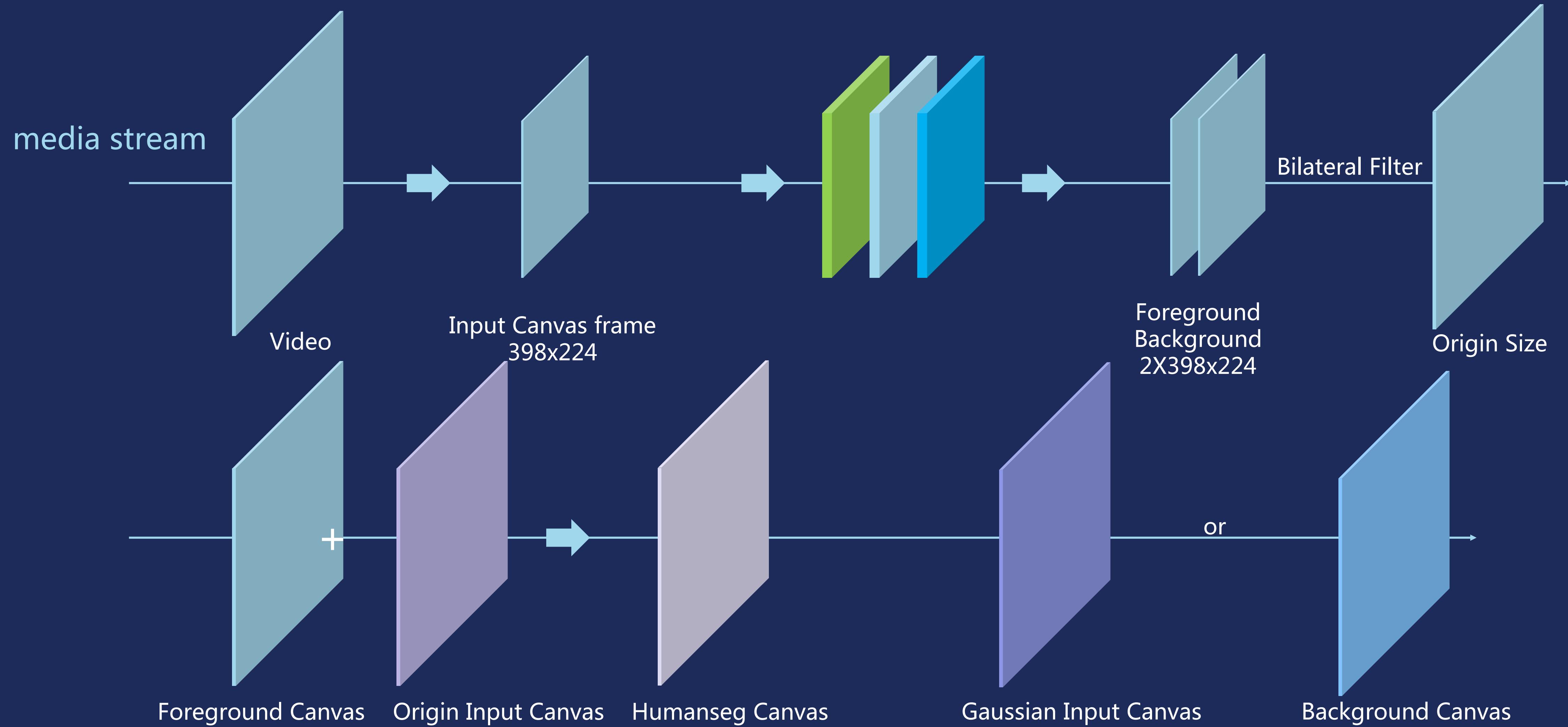
---

- Paddle.js 2.0
- Web 视频会议背景替换
- 开放平台
- 未来展望

# WEB 视频会议背景替换



# 技术方案



# 核心代码

```
import Camera from '@paddlejs-mediapipe/camera';
import * as humanseg from '@paddlejs-models/humanseg/index_shuffle';

const videoCanvas = document.createElement('canvas') as HTMLCanvasElement;
const videoCanvasCtx = videoCanvas.getContext('2d');
await humanseg.load();
const camera = new Camera(video, {
  onFrame: async () => {
    videoCanvas.width = video.width;
    videoCanvas.height = video.height;
    videoCanvasCtx.drawImage(video, 0, 0, video.width, video.height);
    const {
      data
    } = await humanseg.getGrayValue(videoCanvas);
    humanseg.drawHumanSeg(canvas1, data);
  }
});
```

# 模型迭代——PaddleSeg

## 1 数据合成

问题：场景不够丰富，特征场景数据较少

解决方案：数据合成产生大量训练样本

## 2 网络优化

问题：反卷积操作存在棋盘效应

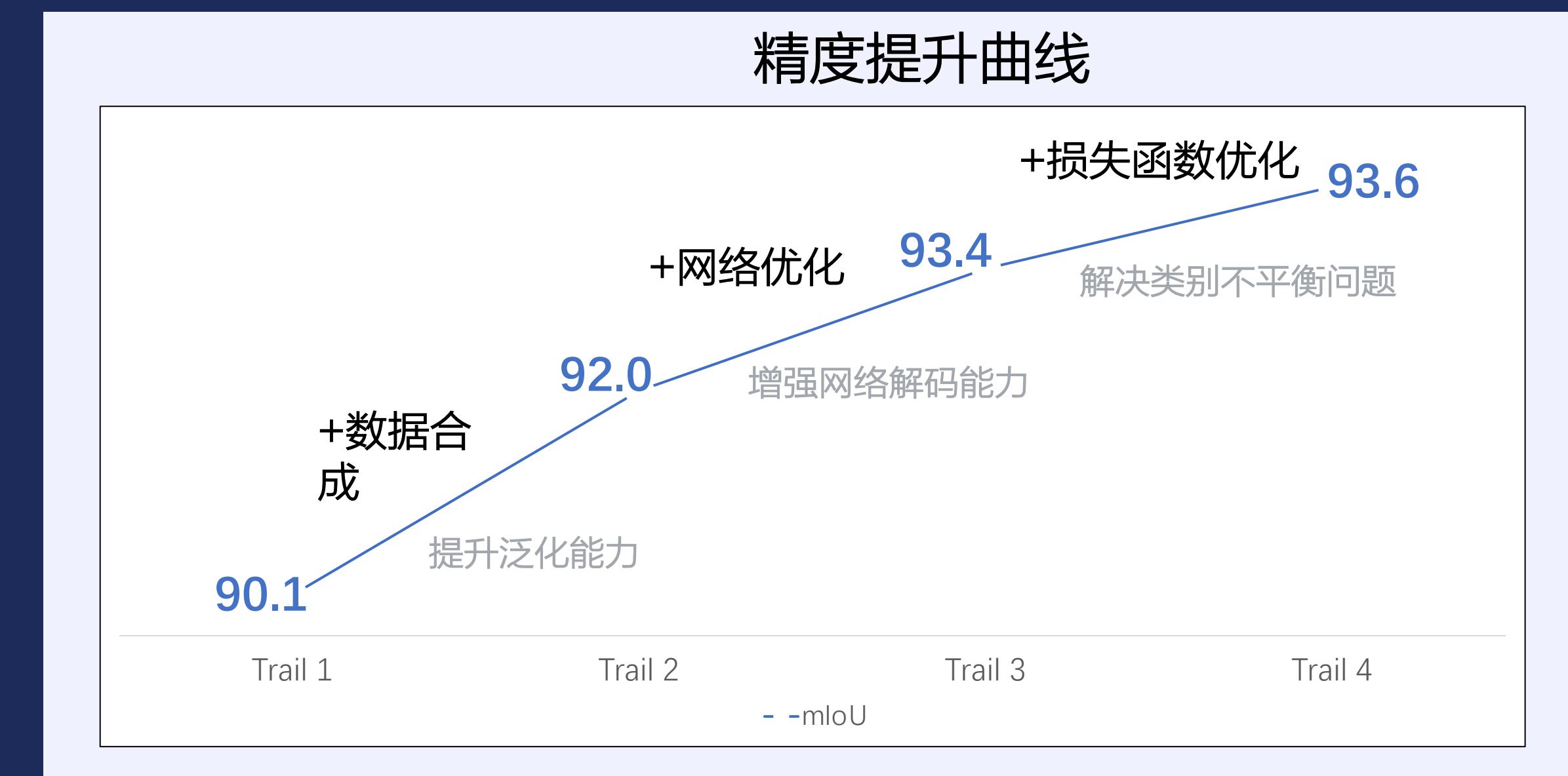
解决方案：使用卷积+双线性插值替代

## 3 损失函数优化

问题：前景、背景类别占比不均衡

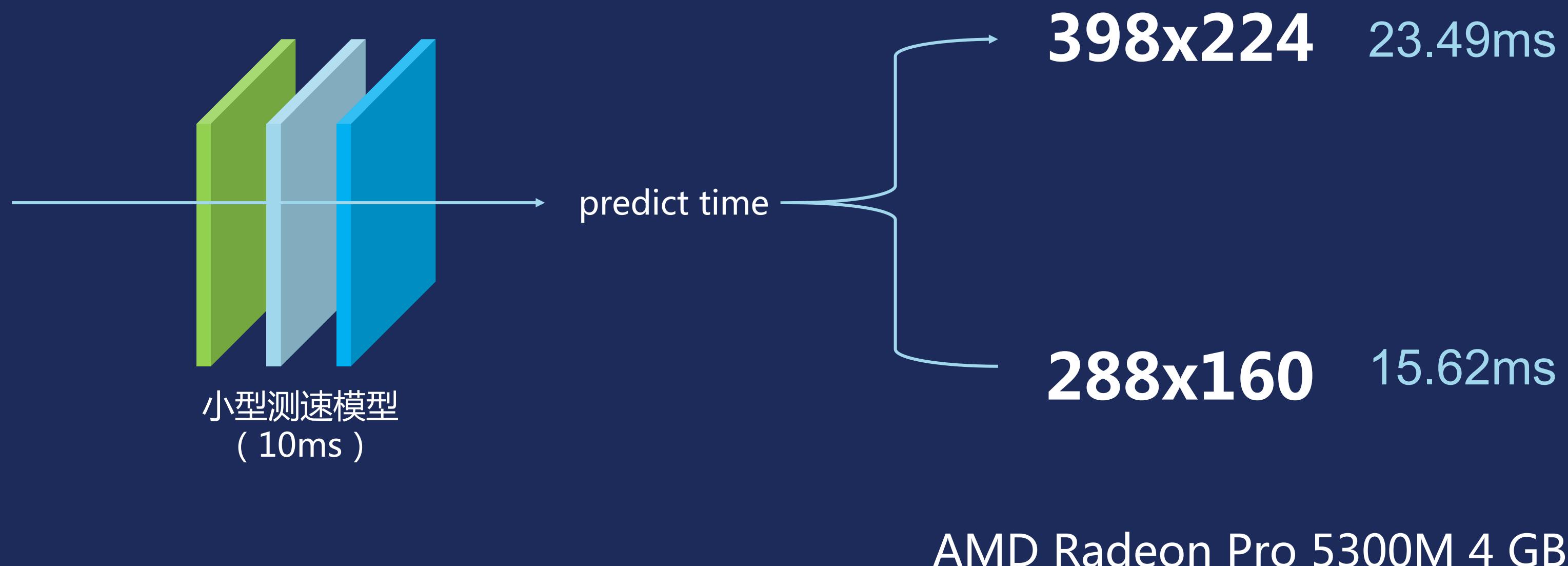
解决方案：采用Lovasz loss监督网络训练

Github Repo:  
<https://github.com/PaddlePaddle/PaddleSeg>



# 模型降级方案

Chrome、Edge、Firefox

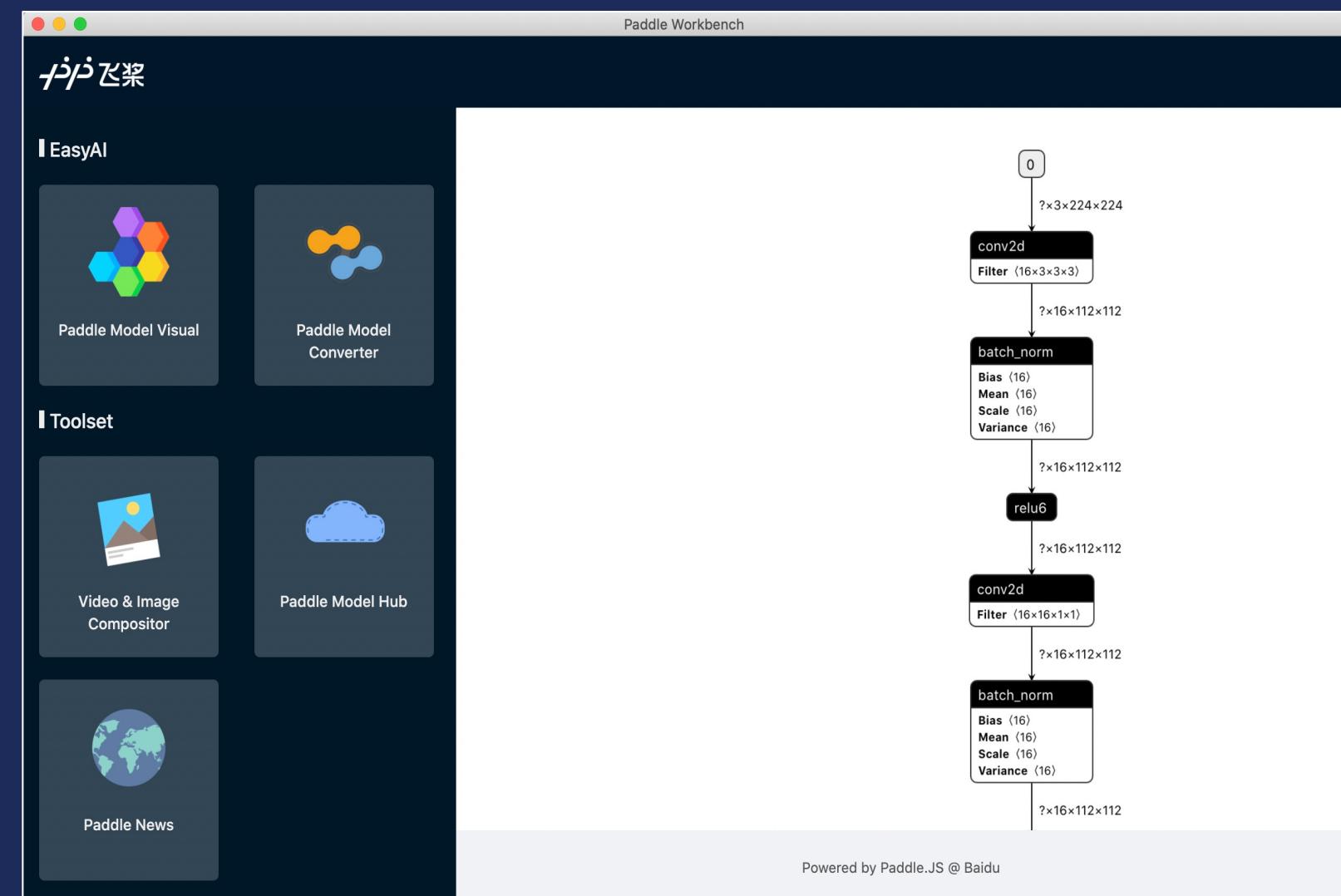
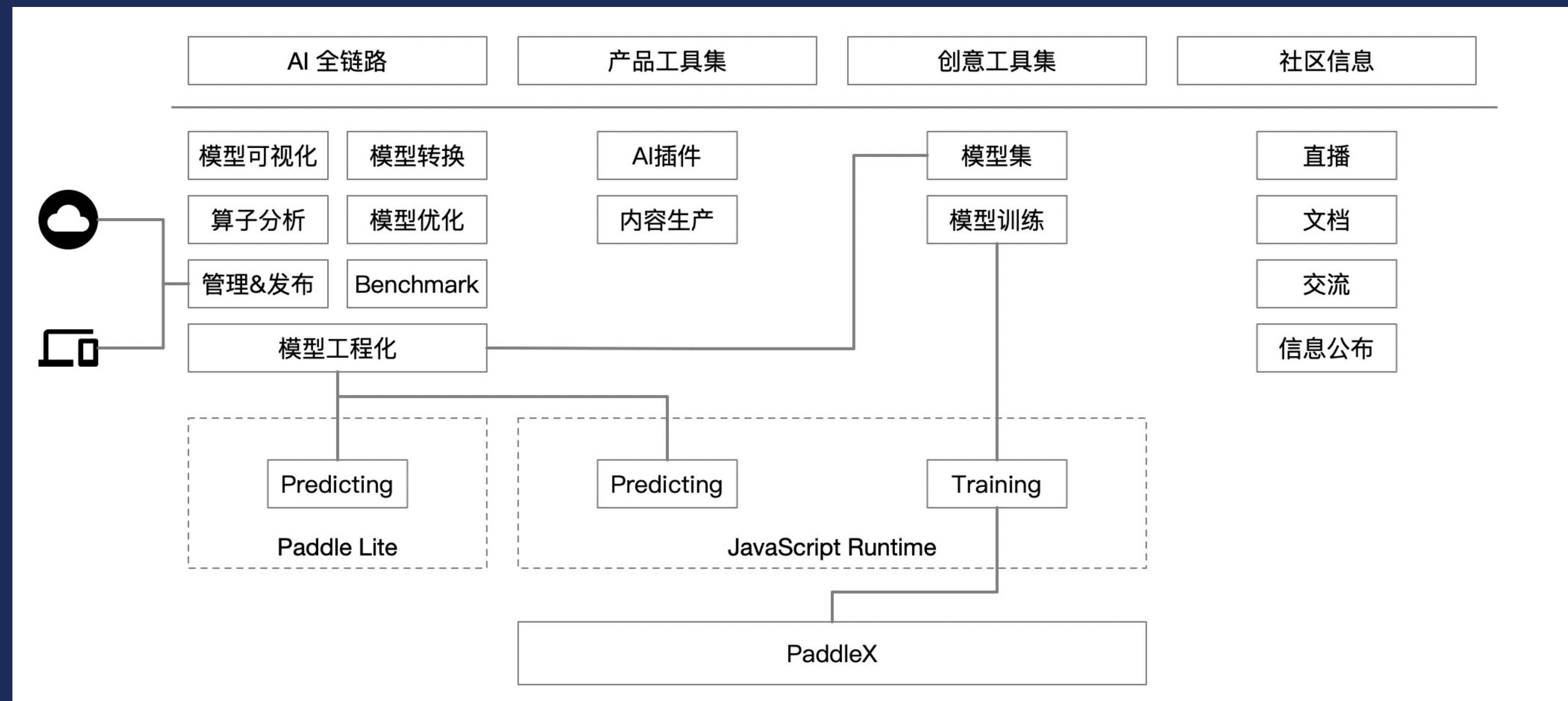


# 大纲

---

- Paddle.js 2.0
- Web 视频会议背景替换
- **开放平台**
- 未来展望

# EASYAI



# 大纲

---

- Paddle.js 2.0
- Web 视频会议背景替换
- 开放平台
- 未来展望

# 未来展望

调优效率

调优方式： 手动 →

AI 编译器

行业解决方案

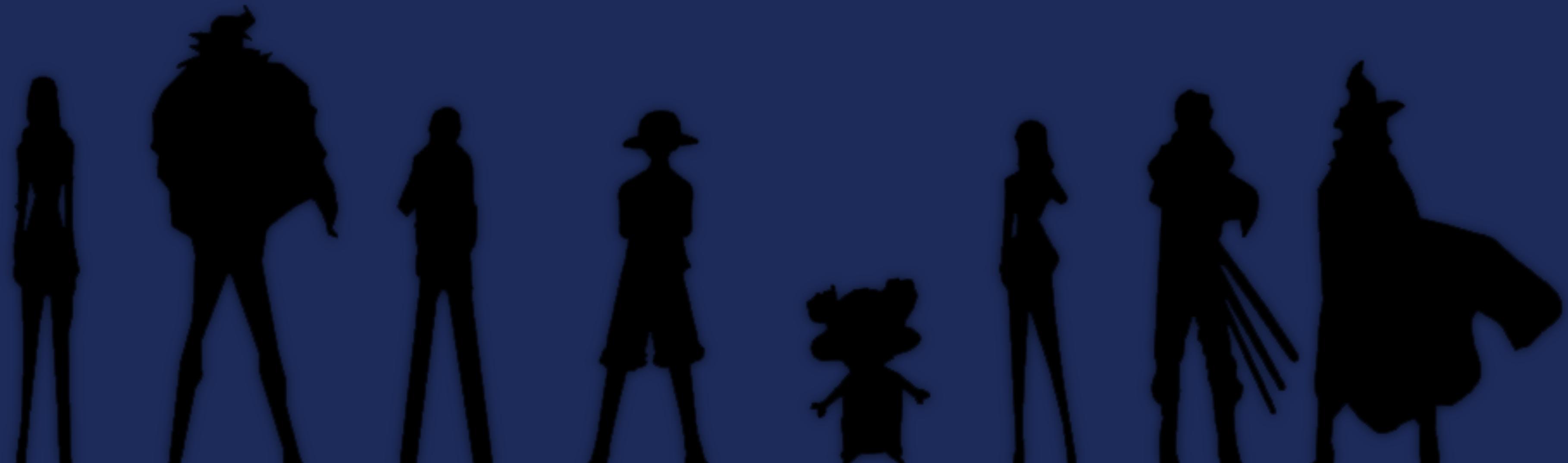
行业智能化解决方案全链路

# Q&A

Paddle.js 微信技术交流群



Paddle.js Github





全球大前端技术大会

# THANKS

InfoQ  
ueue

