

# T/SZSSIA

团 体 标 准

T/SZSSIA XXXX—XXXX

## 基于端边云协同场景的算法模型部署技术要求

Technical requirements of AI algorithm model deployment based on  
end-edge-cloud collaboration scenario

(征求意见稿)

2021-10-13

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX-XX-XX 发布

XXXX-XX-XX 实施

深圳市智慧安防行业协会 发 布

行业协会

深圳市智慧安防行业协会

深圳市智慧安防行业

深圳

深圳市智慧安防行业协会

深圳市智慧安防行业协会

## 目 次

前言 .....	II
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 缩略语 .....	1
5 设计原则 .....	1
5.1 安全性 .....	1
5.2 扩展性 .....	1
5.3 开放性 .....	1
6 系统构成 .....	1
6.1 概述 .....	1
6.2 部署流程 .....	2
7 功能要求 .....	4
7.1 注册与注销 .....	4
7.2 心跳保活 .....	4
7.3 算法模型部署 .....	4
8 接口要求 .....	5
8.1 基本要求 .....	5
8.2 注册请求与响应 .....	5
8.3 心跳保活请求与响应 .....	6
8.4 算法模型部署请求与响应 .....	7
8.5 算法模型部署命令与响应 .....	8
8.6 注销请求与响应 .....	9
8.7 错误码定义 .....	10
9 性能要求 .....	11
9.1 请求/命令超时时间 .....	11
9.2 心跳周期 .....	11
9.3 传输时延 .....	11

## 前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由深圳云天励飞技术股份有限公司提出。

本文件由深圳市智慧安防行业协会归口。

本文件起草单位：

本文件主要起草人：

## 基于端边云协同场景的算法模型部署技术要求

### 1 范围

本文件规定了端边云协同场景的算法模型部署的设计原则、系统构成、功能要求、接口要求和性能要求。

本文件适用于终端和/或边缘设备向云端请求和更新特定的算法模型、云端向终端和/或边缘设备下发并更新特定的算法模型。

### 2 规范性引用文件

本文件没有规范性引用文件。

### 3 术语和定义

下列术语和定义适用于本文件。

#### 3.1 端边云协同 end-edge-cloud collaboration

终端和/或边缘设备与云端协作以满足特定应用场景的需求。

注：在本文件中的特定应用场景为端边云协同完成算法模型的部署。

### 4 缩略语

下列缩略语适用于本文件。

HTTP：超文本传输协议（HyperText Transfer Protocol）

JSON：JS对象简谱（JavaScript Object Notation）

### 5 设计原则

#### 5.1 安全性

系统应具有通过加密传输、身份认证、权限控制等方式实现算法模型部署安全性的措施和方法。

#### 5.2 扩展性

系统的软件接口应具有可扩展性，可满足未来新场景的需求。

#### 5.3 开放性

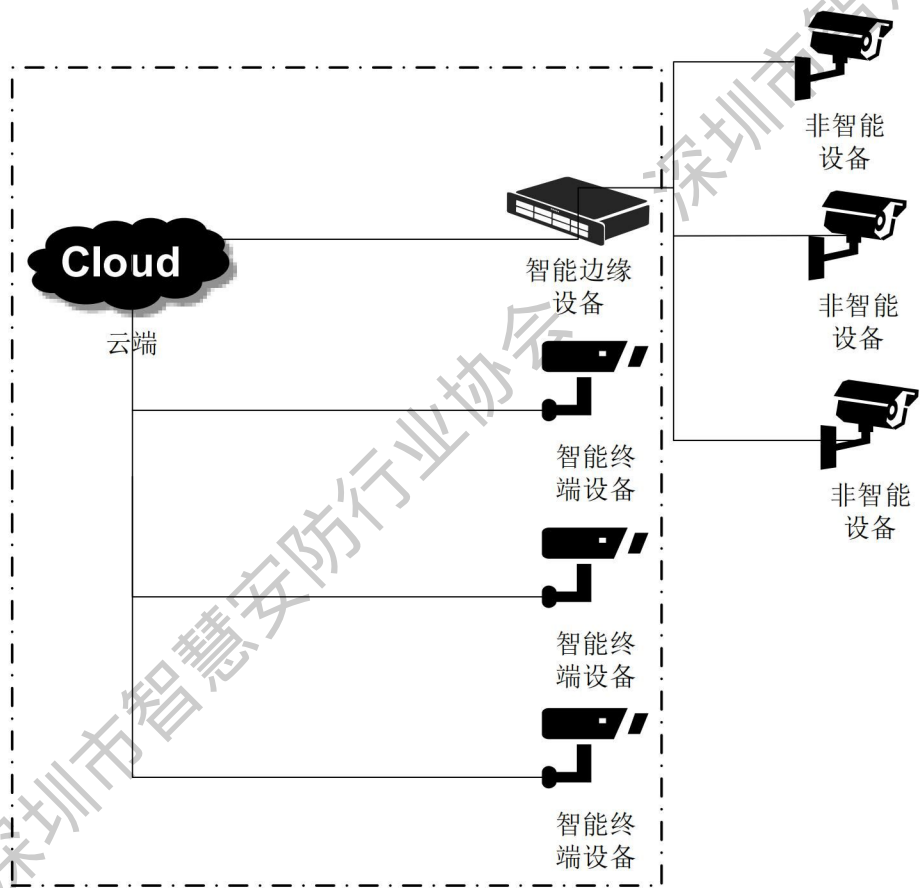
系统的设计应遵循开放性原则，支持对接各种设备，支持二次开发。

### 6 系统构成

#### 6.1 概述

6.1.1 一个典型的端边云协同系统主要由智能终端设备、智能边缘设备、云端等组成，其中智能边缘设备通过网络连接非智能设备。如图 1 所示。

6.1.2 系统的算法模型部署配置功能包括注册与注销、心跳保活、算法模型部署等功能。



说明：

端边云协同场景下的算法模型部署相关流程发生在虚框范围内；云端对算法模型的管理、云端对终端和/或边缘设备的管理以及算法模型在终端和/或边缘设备中的运行和应用不在本文件的描述范围之内。

图1 典型的端边云协同系统构成示意图

6.2 部署流程

6.2.1 端边云协同场景下的算法模型部署流程见图 2。



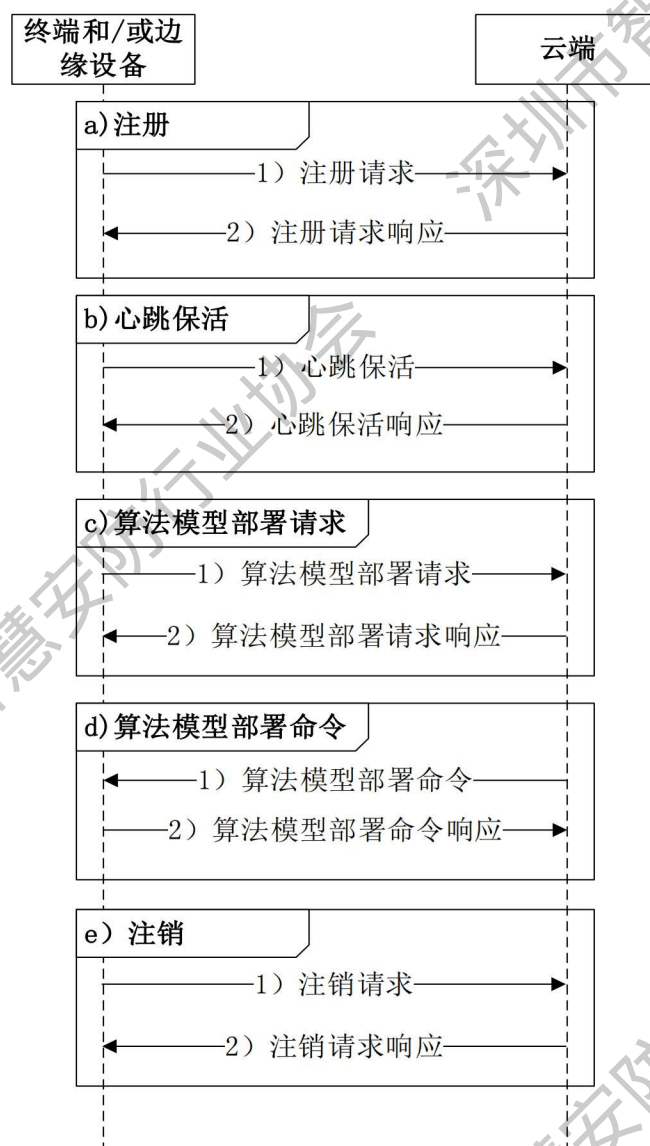


图2 端边云协同场景下的算法模型部署流程

6.2.2 端边云协同场景下的算法模型部署的主要流程包括注册、心跳保活、算法模型部署请求（终端和/或边缘端发起）、算法模型部署命令（云端发起）、注销等过程，符合以下要求：

- a) 注册：
  - 1) 终端和/或边缘设备向云端发起注册请求。云端收到注册请求后应向终端和/或边缘设备发送注册请求响应，回复注册结果；
  - 2) 终端和/或边缘设备在规定的时间内未收到云端的响应，应重新发起注册请求。
- b) 心跳保活：
  - 1) 成功注册的终端和/或边缘设备，应周期性地向云端发送心跳保活消息，云端收到心跳保活消息后，应向终端和/或边缘设备发送心跳保活响应；
  - 2) 终端和/或边缘设备在规定的时间内未收到心跳保活响应，应记为一次心跳超时；
  - 3) 连续发生心跳超时的次数超过规定的次数，终端和/或边缘设备应重新向云端发起注册请求；
  - 4) 云端在规定的时间内未收到心跳保活消息，应认为终端和/或边缘设备处于未注册状态，此时云端收到终端和/或边缘设备的其他消息时均应回复失败，终端和/或边缘设备应重新向云端发起注册请求。

- c) 算法模型部署请求：
  - 1) 终端和/或边缘设备向云端请求部署指定的算法模型，应向云端发送算法模型部署请求；
  - 2) 云端收到算法模型部署请求后向终端和/或边缘设备发送算法模型部署请求响应，拒绝算法模型部署请求或下发请求的算法模型。
- d) 算法模型部署命令：
  - 1) 云端设备要求终端和/或边缘设备部署指定的算法模型，应向终端和/或边缘设备发送算法模型部署命令；
  - 2) 终端和/或边缘设备收到算法模型部署命令后向云端发送算法模型部署命令响应，拒绝算法模型部署命令或者接受指定的算法模型。
- e) 注销：
  - 1) 终端和/或边缘设备向云端发起注销请求；
  - 2) 云端收到注册请求后，向终端和/或边缘设备发送注销请求响应，回复注销结果。

## 7 功能要求

### 7.1 注册与注销

注册与注销应具备以下功能：

- a) 未注册的终端和/或边缘设备应先向云端发起注册请求。注册通过后才可以与云端进行后续的交互；
- b) 云端应对请求注册的终端和/或边缘设备进行合法性和权限校验：
  - 1) 云端校验通过，应向终端和/或边缘设备返回注册成功，并分配唯一的设备标识；
  - 2) 云端检查设备非法，应向终端和/或边缘设备返回注册失败，原因为设备非法；
  - 3) 云端校验设备权限不通过，应向终端和/或边缘设备返回注册失败，原因为鉴权失败；
  - 4) 云端判断设备当前已经注册，应向终端和/或边缘设备返回注册失败，原因为重复注册。
- c) 终端和/或边缘设备超过 30s 未收到注册请求响应，应重新发起注册请求；
- d) 终端和/或边缘设备收到注册失败的响应后，应等待 300s 重新发起注册请求；
- e) 已注册的终端和/或边缘设备注销时应向云端发起注销请求。注销请求每隔 30s 发送一次，连续发送 3 次，无论期间是否收到注销请求响应，注销流程均正常结束；
- f) 云端每次收到终端和/或边缘设备的注销请求后，直接回复响应消息，结果为注销成功。

### 7.2 心跳保活

心跳保活应具备以下功能：

- a) 终端和/或边缘设备在收到注册成功消息后，应立即发送心跳保活消息；
- b) 云端收到心跳保活消息后，应回复心跳保活响应；
- c) 终端和/或边缘设备在心跳保活消息发出 30s 内未收到心跳保活响应，记录一次心跳超时事件。如果连续发生 3 次心跳超时事件，终端和/或边缘设备应重新向云端发起注册请求；
- d) 云端向终端和/或边缘设备回复注册成功后 90s 内未收到心跳保活消息，应认为终端和/或边缘设备处于未注册状态。此时云端收到终端和/或边缘设备的其他消息时均需要回复失败，原因为设备未注册。

### 7.3 算法模型部署

算法模型部署应具备以下功能：

- a) 终端和/或边缘设备根据自身的硬件类型、设备型号和支持的算法模型请求指定的算法模型，应向云端发送算法模型部署请求：
  - 1) 云端没有请求的算法模型时，应回复请求失败，原因为模型不存在；
  - 2) 云端有请求的算法模型时，应回复请求成功，同时携带算法模型下载地址。
- b) 云端向终端和/或边缘设备推送算法模型，应发送模型部署命令：
  - 1) 终端和/或边缘设备判断命令中的算法模型与设备不兼容，应向云端回复失败，原因为模型不兼容；



- 2) 终端和/或边缘设备判断命令中的算法模型版本不高于设备本地已经部署的算法模型版本，应向云端回复失败，原因为不需要部署；
- 3) 终端和/或边缘设备判断命令中的算法模型满足部署条件，应向云端回复部署成功，同时下载并运行命令中的模型。

## 8 接口要求

### 8.1 基本要求

接口的基本要求如下：

- a) 接口协议：HTTP 协议；
- b) 数据格式：JSON 格式。

### 8.2 注册请求与响应

注册请求与响应符合以下要求：

- a) 信令流程见图 3，步骤如下：
  - 1) 终端和/或边缘设备向云端发起注册请求 RegisterRequest；
  - 2) 云端向终端和/或边缘设备回复注册响应 RegisterResponse。



图3 注册请求与响应信令流程

- b) 消息格式见表 1。

表1 注册请求与响应消息格式

URI	/api/ModelDeployment/RegisterRequest			
功能	终端和/或边缘设备注册			
Content-Type	application/JSON			
请求消息字段说明				
名称	说明	数据类型	是否必选	备注
DeviceID	设备序列号	String	是	
请求消息示例				
RegisterRequest	{ "DeviceID": "VendorXX1234567890" }			
返回消息字段说明				
名称	说明	数据类型	是否必选	备注

T/SZSSIAXXXX—XXXX

Code	错误码	Int	是	错误码含义详见 8.7
DeviceID	设备序列号	String	是	
UuID	设备唯一注册标识	String	否	Code 为 200 时必填
返回消息示例				
RegisterResponse	{ "Code":200, "DeviceID":"VendorXX1234567890", "Uuid":"44031000000000000001" }			

8.3 心跳保活请求与响应

心跳保活请求与响应符合以下要求：

- a) 信令流程见图 4，步骤如下：
- 1) 终端和/或边缘设备向云端发送心跳保活消息 HeartBeat；
  - 2) 云端向终端和/或边缘设备回复心跳保活响应 HeartBeatResponse。

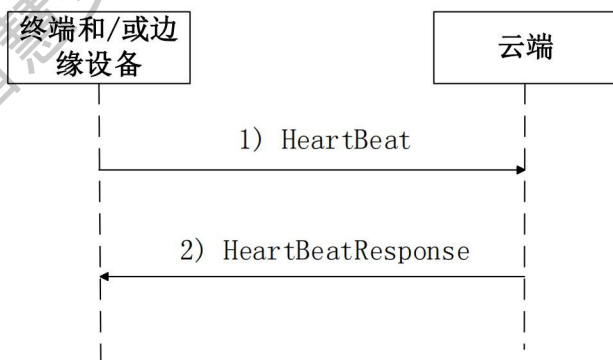


图4 心跳保活请求与响应信令流程

- b) 消息格式见表 2。

表2 心跳保活请求与响应消息格式

URI	/api/ModelDeployment/HeartBeat			
功能	心跳保活			
Content-Type	application/JSON			
请求消息字段说明				
名称	说明	数据类型	是否必选	备注
UuID	设备唯一注册标识	String	是	
请求消息示例				
HeartBeat	{ "UuID": "440310000000000000001" }			
返回消息字段说明				
名称	说明	数据类型	是否必选	备注
Code	错误码	Int	是	错误码含义详见 8.7

UuID	设备唯一注册标识	String	是	
返回消息示例				
HeartBeatResponse	<pre>{   "Code":200,   "UuID":"44031000000000000001" }</pre>			

#### 8.4 算法模型部署请求与响应

算法模型部署请求与响应符合以下要求：

- a) 信令流程见图 5，步骤如下：
- 1) 终端和/或边缘设备向云端发送算法模型部署请求 ModelDeployRequest；
  - 2) 云端向终端和/或边缘设备回复算法模型部署响应 ModelDeployResponse。

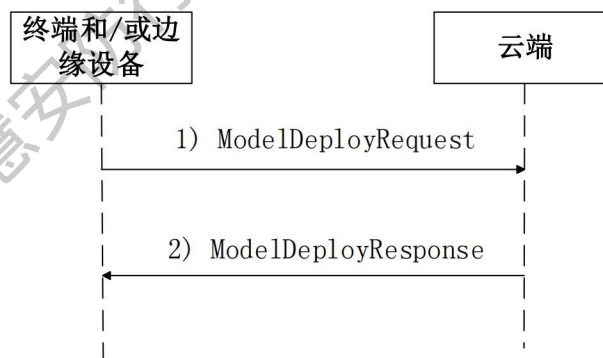


图5 算法模型部署请求与响应信令流程

- b) 消息格式见表 3。

表3 算法模型部署请求与响应消息格式

URI			/api/ModelDeployment/ModelDeployRequest			
功能			终端和/或边缘设备请求部署模型			
Content-Type			application/JSON			
请求消息字段说明						
名称			说明	数据类型	是否必选	备注
UuID			设备唯一注册标识	String	是	
ReqInfo			模型请求信息	ReqInfo	是	
	Vendor		设备厂商	String	是	
	DeviceType		设备类型	String	是	
	DeviceModel		设备型号	String		
	ModelReqList		模型请求列表	List<ModelReqList>	是	
		Name	模型名称	String	是	
		Version	模型版本号	String	是	latest 表示请求最新版本

请求消息示例						
ModelDeployRequest			{ "UUID ":"44031000000000000001", "ReqInfo": { "Vendor": "vendor name", "DeviceType": "IPC", "DeviceModel": "IPC-N11", "ModelReqList": [ { "Name": "FaceDet", "Version": "1.0.0", }] } }			
返回消息字段说明						
名称			说明	数据类型	是否必选	备注
UUID			设备唯一注册标识	String	否	
ModelList			模型列表	List<ModelList>		与请求消息中的ModelReqList一一对应
	ModelUrl		模型文件下载链接	String		
	Code		错误码	Int		错误码含义详见 8.7
返回消息示例						
ModelDeployResponse			{ "UUID ":"44031000000000000001", "ModelList ": [{ "ModelUrl": "/model/ipc/FaceDet-1.0.0", "Code": 200 }] }			

### 8.5 算法模型部署命令与响应

算法模型部署命令与响应符合以下要求：

- a) 信令流程见图 6，步骤如下：
- 1) 终端和/或边缘设备向云端发送算法模型部署命令 ModelDeployCommand；
  - 2) 云端向终端和/或边缘设备回复算法模型部署响应 ModelDeployResponse。

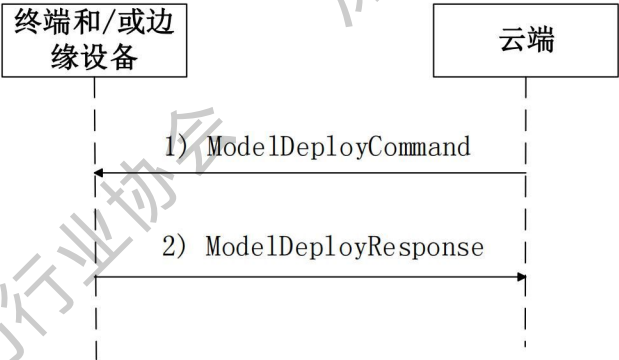


图6 算法模型部署命令与响应信令流程

b) 消息格式见表 4。

表4 算法模型部署命令与响应消息格式

URI			/api/ModelDeployment/ModelDeployCommand			
功能			云端向终端和/或边缘设备发送模型部署命令			
Content-Type			application/JSON			
请求消息字段说明						
名称			说明	数据类型	是否必选	备注
UuID			设备唯一注册标识	String	是	
ModelCmdList			模型部署命令列表	List<ModelCmdList>		
	Vendor		设备厂商	String	是	
	DeviceType		设备类型	String	是	
	DeviceModel		设备型号	String	是	
	Name		模型名称	String	是	
	Version		模型版本号	String	是	
	ModelUrl		模型文件下载链接	String	是	
请求消息示例						
ModelDeployCommand			{ "UuID": "44031000000000000001", "ModelCmdList": [{ "Vendor": "vendor name", "DeviceType": "IPC", "DeviceModel": "IPC-N11", "Name": "FaceDet", "Version": "1.0.0", "ModelUrl": "/model/ipc/FaceDet-1.0.0" }] }			
返回消息字段说明						
名称			说明	数据类型	是否必选	备注
UuID			设备唯一注册标识	String	是	
ResultList				List<ResultList>		与命令消息中的ModelCmdList 一一对应
	Code		错误码		是	错误码含义详见8.7
返回消息示例						
ModelDeployResponse			{ "Uuid": "44031000000000000001" "ResultList": [0] }			

## 8.6 注销请求与响应



注销请求与响应符合以下要求：

- a) 信令流程见图 7，步骤如下：
- 1) 终端和/或边缘设备向云端发起注销请求 UnregisterRequest；
  - 2) 云端向终端和/或边缘设备回复注销响应 UnregisterResponse。



图7 注销请求与响应信令流程

- b) 消息格式见表 5。

表5 注销请求与响应消息格式

URI	/api/ModelDeployment/UnregisterRequest			
功能	终端和/或边缘设备注销			
Content-Type	application/JSON			
请求消息字段说明				
名称	说明	数据类型	是否必选	备注
UUID	设备唯一注册标识	String	是	
请求消息示例				
UnregisterRequest	{ "UUID ":"4403100000000000000001" }			
返回消息字段说明				
名称	说明	数据类型	是否必选	备注
Code	错误码	Int	是	错误码含义详见 8.7
UUID	设备唯一注册标识	String	是	
返回消息示例				
UnregisterResponse	{ "Code":200, "UUID ":"4403100000000000000001" }			

## 8.7 错误码定义

错误码定义见表6。

表6 错误码定义

错误代码	错误说明
200	成功
201	非法设备
202	鉴权失败
203	重复注册
204	未注册
301	请求的算法模型文件不存在
302	指定的模型无需部署
401	服务器内部错误
502	下载模型失败
503	模型完整性校验失败
504	模型部署失败

## 9 性能要求

### 9.1 请求/命令超时时间

请求和命令消息发送后，发送超过30s未收到对方的响应，则认为消息发送超时，可以尝试重发该消息。

### 9.2 心跳周期

心跳交互的性能满足以下要求：

- 终端和/或边缘设备收到注册成功的响应后，应立刻发送心跳保活消息；
- 终端和/或边缘设备发送心跳保活消息的间隔宜为 30s；
- 终端和/或边缘设备连续 3 次心跳超时后，需要重新发送注册请求；
- 云端在向终端和/或边缘设备回复注册成功 90s 内未收到心跳保活消息，则认为终端和/或者边缘设备处于未注册状态。

### 9.3 传输时延

在设备和网络环境正常的情况下，各个交互消息的传输时延宜不超过1s。