



# AI 视觉训练 数据智能生成平台

为国防领域打造的合成数据生成与自动标注解决方案

---

产品白皮书

北京数字冰雹信息技术股份有限公司

## 1. 产品概述

### 1.1 平台定位

AI 视觉训练数据智能生成平台 是一款专为国防领域设计的合成数据生成与自动标注解决方案。该平台通过智能体驱动数字孪生引擎，实现多模态、高保真视觉训练数据的智能化与自动化生成，针对战场环境中数据获取成本高、风险大的挑战提供解决方案。

平台融合了先进的数字孪生场景构建能力和智能体技术，使用户无需专业训练即可快速创建涵盖城市、野外、海洋、空间及复杂天气光照条件的虚拟作战环境，并批量产出包含光电、红外等传感器数据，以及目标深度、材质、语义分割等高精度多模态数据，同时自动完成像素级标注。

平台旨在为自主决策系统提供可靠、可控的数据基石，加速国防 AI 系统的训练、测试与验证流程。

### 1.2 解决的核心问题

在国防 AI 系统开发中，真实战场数据的获取往往面临高昂成本、安全风险和操作限制。传统数据采集方法难以覆盖极端场景，如复杂天气、光照变化或罕见威胁情境，导致模型泛化能力不足。

本平台通过智能体与数字孪生技术的深度融合，解决了数据稀缺性、标注效率低以及模型鲁棒性差等问题。用户可基于平台快速构建逼真的数字孪生场景，并利用智能体自动化生成多样化的训练数据，从而降低对真实数据的依赖，提升模型在真实环境中的表现。这不仅缩短了开发周期，还确保了数据的安全性与可控性，为国防应用提供了高效、经济的数据解决方案。

## 2. 产品功能

### 2.1 智能体驱动的数字孪生场景构建

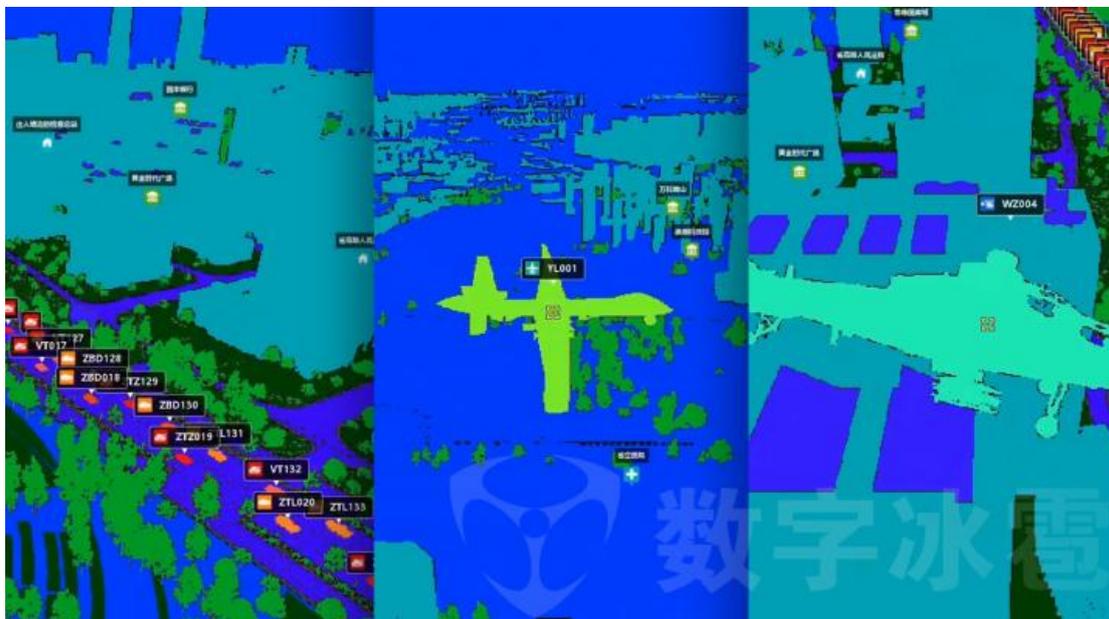
平台通过智能体技术实现数字孪生场景的自动化构建与优化。用户可利用可视化编辑器，以拖拽方式快速设计虚拟环境，无需专业建模知识。智能体基于语义向量和知识库检索，自动调用预置资源库（如模型资产、效果模板和 GIS 数据），生成涵盖城市、野外、海洋等多样场景。该功能支持大规模环境生成，包括自动路网规划、建筑布局 and 地形模拟，并能根据用户需求动态调整场景参数，如光照、天气和时间周期。这降低了场景构建的门槛，使非技术用户也能在短时间内创建高保真作战环境，为后续数据生成奠定基础。



### 2.2 多模态视觉训练数据生成

平台能够批量生成包含光电、红外和雷达等多传感器数据的多模态视觉训练数据集。基于物理确定的渲染引擎，平台精确模拟传感器在动态条件下的行为，输出高保真图像和地面实况数据。用户可自定义数据格式和分辨率，生成包括彩色图像、深度图、法线向量和光谱辐射等输出类型。

该功能支持多种环境条件，如雨、雪、雾和眩光，确保数据覆盖真实世界中的边缘案例。通过程序化生成，平台可在几分钟内产出大量统计平衡的数据集，满足不同 AI 模型的训练需求，提升模型在复杂场景中的感知精度。



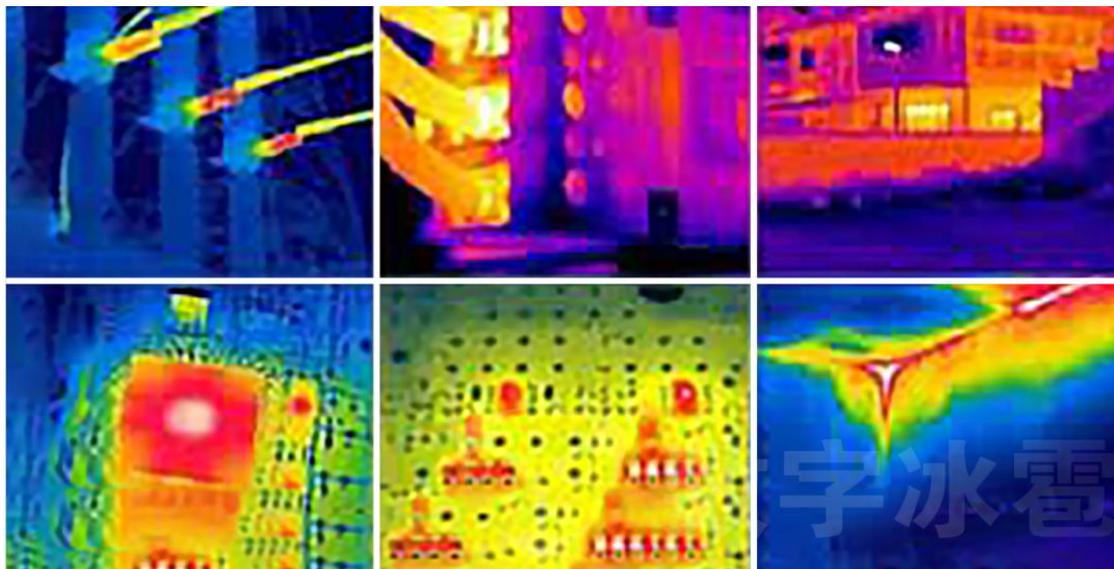
## 2.3 自动标注与地面实况生成

平台提供全自动的像素级标注服务，生成包括语义分割、实例分割、材质分类和运动矢量等高精度元数据。利用智能体驱动的标注引擎，平台根据用户定义的本体和分类法，自动为每帧数据添加标签，无需人工干预。标注输出支持多种格式，如 8 位图像用于标签和实例，32 位文件用于深度和 3D 位置数据。该功能不仅提高了标注效率，还确保了数据的一致性和准确性，减少人为错误。用户可通过参数编辑快速迭代标注规范，使生成的数据集直接集成到机器学习管道中，加速模型验证与部署。



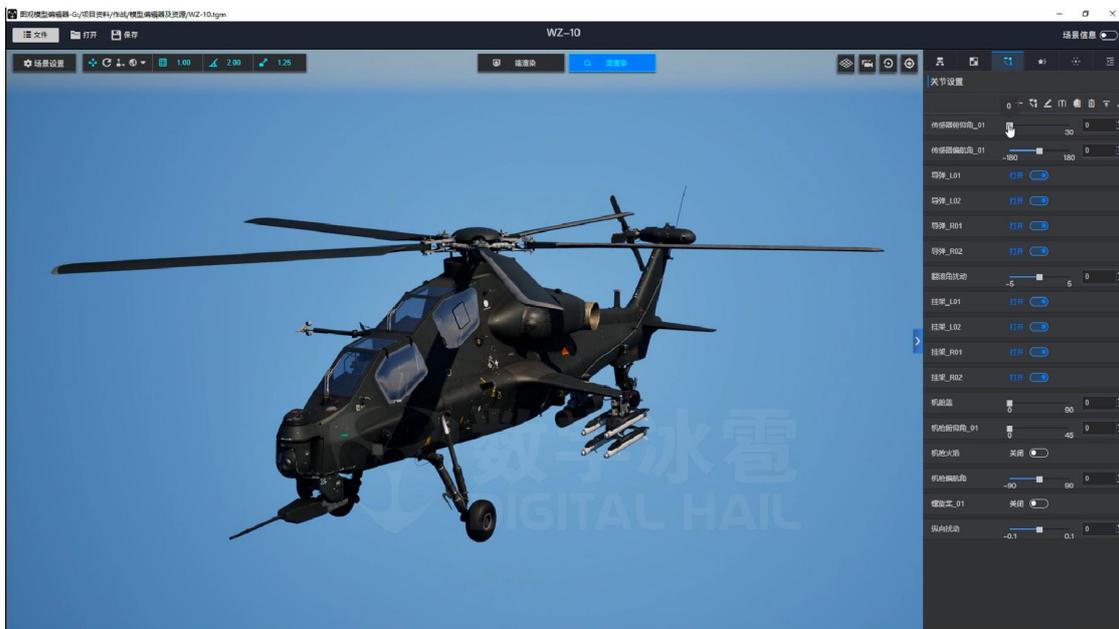
## 2.4 多传感器仿真与集成

平台支持红外、雷达和热成像传感器的物理精确仿真，模拟真实传感器的响应和噪声特性。通过多光谱引擎，平台计算每束光束的光谱辐射度，生成传感器特定的输出，如原始图像和辐射度数据。该功能允许用户配置复杂的传感器阵列，并模拟不同环境下的性能变化，例如低光照或恶劣天气。集成后的数据可直接用于训练多传感器融合模型，增强系统在真实任务中的泛化能力。此外，平台提供统一的 API 接口，便于与现有测试框架集成，实现端到端的传感器数据验证。



## 2.5 可视化编辑与低代码开发

平台提供直观的可视化编辑工具，支持零代码和低代码应用开发。用户可以通过拖拽方式配置场景、交互逻辑和数据源，无需编程技能即可构建数字孪生应用。编辑器内置丰富的组件库，包括图表、控件和图层，支持跨数据源联动和参数化交互。对于高级用户，平台提供基于 JavaScript 的统一开发 API，允许自定义控制场景对象、特效和 GIS 数据。该功能兼容主流前端框架，如 Vue 和 React，并支持端渲染和流渲染模式的无缝切换。这大幅降低了开发门槛，使团队能够快速原型化和部署 AI 训练环境，提高整体生产效率。



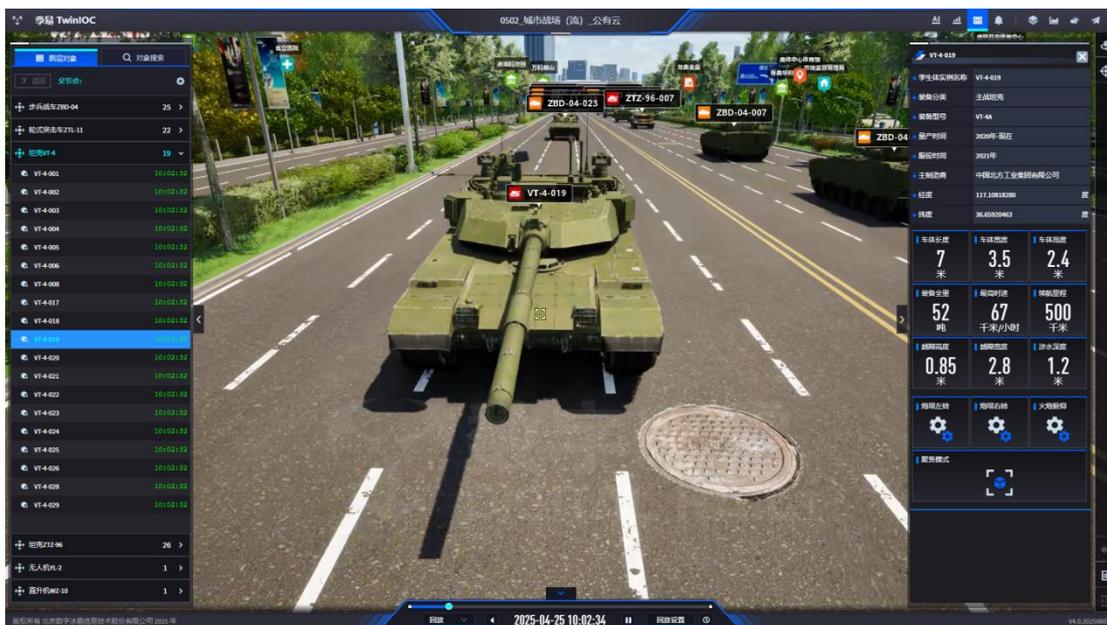
## 2.6 企业级安全与权限管理

平台具备精细化的权限控制和安全管理机制，确保数据在生成、存储和访问过程中的安全性与合规性。通过角色基础的访问控制，管理员可严格管理用户对场景、数据和智能体的操作权限，防止未授权访问。平台支持私有化部署，数据完全存储于内网环境，满足国防领域对数据保密性的高要求。

## 3. 核心技术

### 3.1 基于物理确定性的渲染引擎

平台采用专有的光谱路径追踪引擎，以物理精度模拟光线传输和材质交互。该引擎计算每束光束的光谱辐射度，生成高度逼真的图像和传感器数据，确保输出与真实世界的一致性。通过支持多层材质纹理和 PBR 物理材质渲染，引擎能够还原场景的细微细节，如反射率和粗糙度，为多模态数据生成提供可靠基础。这项技术显著缩小了仿真与真实数据之间的领域差距，提升 AI 模型在部署中的性能。



### 3.2 多光谱传感器仿真技术

平台的多光谱仿真能力可精确模拟多种传感器，包括 RGB-IR 相机、激光雷达和雷达，输出传感器特定的数据流。技术基于物理模型，模拟传感器在动态环境中的行为，如噪声、畸变和响应延迟。通过生成辐射度和光谱保真度数据，平台确保输出与目标传感器的特性匹配，支持高级应用如目标检测和地形分割。该技术还允许用户自定义传感器配置，适应不同任务需求，增强数据的实用性和泛

化性。



### 3.3 程序化场景生成与域随机化

利用基于分布的域随机化，平台实现大规模场景的程序化生成，自动创建多样化的虚拟环境。技术通过随机化参数如光照、天气和物体布局，生成统计平衡的数据集，覆盖常见和边缘案例。用户可定义生成规则和本体，确保数据与项目分类法一致。该技术不仅加快了场景构建速度，还提高了数据集的多样性，减少模型过拟合风险，使其在真实环境中更具鲁棒性。

### 3.4 智能体协同与任务编排

平台集成智能体技术，通过多智能体协同机制自动化复杂任务，如场景构建和数据生成。智能体基于可视化编辑器进行任务编排，支持条件分支和工具调用，无需编码即可设计 workflow。技术还包含语义向量和知识库检索，使智能体能够访问客户私有数据，做出精准决策。通过类即时通讯的会话服务，多个智能体可协作处理任务链，提升整体效率。该核心技术降低了 AI 应用的门槛，实现人机协同的智能化操作。



### 3.5 云原生可扩展架构

平台采用云原生设计,支持弹性扩展,能够处理生产级规模的数据生成任务。通过云引擎,平台分布式管理场景渲染、模拟和存储,快速交付数百万标注样本。架构支持多机集群和负载均衡,确保高并发访问下的性能稳定。用户可根据需求动态调整资源,避免瓶颈问题。该技术使平台能够适应大规模部署,满足企业不断增长的数据需求,同时降低成本。



## 4. 产品优势

### 4.1 降低数据获取成本与风险

平台通过合成数据生成，降低真实数据采集的成本与安全风险。用户可在虚拟环境中模拟极端场景，如战场废墟或恶劣天气，无需实地部署。这大幅减少了物资和人力投入，同时确保数据生成过程的安全可控，为国防应用提供经济高效的解决方案。

### 4.2 提升模型训练效率与泛化性

平台能够快速生成大量多样化的训练数据，覆盖广泛的环境条件和边缘案例，加速模型训练周期。通过多模态和自动标注功能，数据直接集成到机器学习管道，减少预处理时间。生成的数据集经过统计平衡，增强模型在真实世界中的泛化能力，提高 AI 系统的整体性能。

### 4.3 高保真与多模态数据支持

基于物理的渲染和传感器仿真技术，平台输出高保真数据，精确模拟真实场景的视觉和物理特性。支持多传感器输出，如红外、激光雷达和雷达，确保数据全面性。这使模型能够学习复杂环境中的细微特征，提升感知精度，适用于高要求的国防任务。

### 4.4 低门槛与高灵活性

平台的可视化编辑和低代码工具使非技术用户也能轻松操作，快速构建场景和应用。同时，API 和定制选项为开发者提供灵活性，适应复杂需求。这种结合确保了平台既能用于快速原型验证，也能支持核心系统开发，满足多样化的用户需求。

## 4.5 企业级安全与可靠性

平台的设计以安全为先，提供私有化部署、细粒度权限控制，确保全生命周期安全。智能体运行在隔离环境中，防止数据泄露。这使平台符合国防领域的严格合规要求，为用户提供可信赖的数据管理基础。

## 5. 客户价值

### 5.1 加速 AI 系统开发周期

平台通过自动化数据生成和标注，缩短模型训练和测试的时间。用户可在几分钟内产出定制数据集，快速迭代模型设计，从而加快产品上市速度。这在快速演变的国防环境中尤为重要，帮助用户及时响应威胁和机遇。

### 5.2 提升决策精准度与态势感知

平台生成的高精度数据使 AI 模型能够更准确地识别目标、分析地形和预测行为。通过多传感器仿真，系统获得全面的环境信息，增强态势感知能力。这支持指挥系统做出基于数据的决策，有助于提高任务成功率。

### 5.3 优化资源分配与人力成本

平台自动化了繁琐的数据处理任务，减少对专业标注人员的依赖，释放人力资源用于高价值工作。例如，团队可将精力集中在策略开发和模型优化上，提升整体效率并降低运营成本。

## 5.4 增强系统鲁棒性与安全性

通过覆盖边缘案例和极端场景，平台训练出的模型在真实环境中表现更稳定，减少故障风险。同时，私有化部署和安全控制确保数据不被泄露，维护系统的整体安全性，适用于敏感任务。

## 5.5 推动业务创新与核心竞争力

低门槛的开发工具使业务团队能快速实验新想法，如自主无人机或智能监控系统，加速创新闭环。这帮助用户在竞争中抢占先机，构建独特的 AI 驱动能力，提升长期竞争力。

## 6 总结

AI 视觉训练数据智能生成平台集成了智能体技术与数字孪生能力，为国防 AI 系统提供了一套完整、高效的合成数据解决方案。通过自动化场景构建、多模态数据生成和精确标注，平台解决了数据获取的痛点，显著提升了模型训练的效率和可靠性。

平台为客户带来加速开发、优化成本和增强创新等多重价值，是推进自主系统发展的关键工具。选择本平台，意味着拥抱智能化未来，以可靠数据驱动更安全的 AI 应用。