

# 메타버스 내 원격 부동산 중계 시스템을 위한 부동산 매물 영상 내 민감정보 삭제 기술

\*김준호, \*김진홍, \*강병준, \*최재원, \*김지훈, \*강동우  
\*홍익대학교 전자·전기공학부  
\*boy814@naver.com

## Privacy-preserving Proptech using Domain Adaptation in Metaverse

\*Junho Kim \*Jinhong Kim \*Byeongjun Kang \*Jaewon Choi \*Jihoon Kim \*Dongwoo  
Kang

\*Department of Electronic and Electrical Engineering, Hongik University

### 요 약

본 논문은 메타버스 등 인공지능 연계 증강/가상현실 부동산 중계 플랫폼에서 부동산 영상 기반 매물 소개 시스템 구축에서 사생활 및 개인정보가 영상에 담기게 될 수 있는 위험이 존재하기에 부동산 영상 내의 개인정보 및 민감 정보를 인공지능 기술을 기반으로 검출하여 삭제해주고 복원해주는 인공지능 기술 연구개발을 목표로 하였다. 한국형 부동산 내 민감 object 를 정의하고, 최신 인공지능 딥러닝 기술 기반 민감 object detection 알고리즘을 연구 개발하며, 영상에서 삭제된 부분은 인공지능 기술을 기반으로 물체가 없는 실제 공간영상으로 복원해주는 영상복원 기술도 연구 개발하였다. 한국형 부동산 환경 (영상 촬영 조도, 디스플레이 스타일, 주변 가구 배치 등)에 맞는 인공지능 모델 구축을 위하여, 자체적으로 한국 영상 database 구축 및 Transfer learning for target domain adaptation 을 진행하였다. 제안된 알고리즘은 일반적인 환경에서 98%의 정확도와 challenge 환경에서 (occlusion 빛 반사, 저조도 등) 81%의 정확도를 보였다. 본 기술은 Proptech 분야에서 주목받고 있는 메타버스 기반 온라인 중계 서비스 기술을 활성화하기 위하여 기획되었으며, 특히 메타버스 부동산 중계 플랫폼의 활성화를 위하여 사생활 보호 측면에서 필요한 중요 기술을 인공지능 기술을 활용하여 연구 개발하였다.

### 1. Introduction

메타버스 플랫폼에서 인공지능 연계한 증강현실 부동산 중계 플랫폼에서 프롭테크 (Proptech)는 부동산 (Property)와 기술 (Technology)의 합성어로 4 차산업혁명 관련 기술의 발전과 코로나 19 에 따른 비대면 기술에 대한 요구로 인해 관련 기업이 많은 주목을 받고 있다 [1].

미국의 프롭테크 기업인 질로우(Zillow)는 빅데이터 및 머신러닝 기술을 부동산 중개업에 적용하여 2020 년도에 2 억 3600 만명이 넘는 이용자를 유치하였으며 올해 5 월에는 1.22B 에 달하는 매출액을 보여주며 엄청난 성장폭을 보였다 [2]. 이러한

세계적인 추세에도 불구하고 국내에서는 아직 대표적인 기술 기반 프롭테크 기업이 존재하지 않으며, 본 논문에서는 프롭테크 분야에서 주목받고 있는 가상현실 및 증강현실 기반 온라인 중개 서비스 기술을 개발하기 위하여 기획되었으며 이에 따른 문제에 대한 해결책을 제시한다는 목표를 가지고 연구하였다.

영상 기반 매물 소개 시스템 구축을 통해 거주인 혹은 중개사가 자택을 동영상 촬영 후 업로드 혹은 실시간 촬영을 통해 매수/임차인에게 중개하게 되면 사생활 및 개인정보가 영상에 담기게 될 수 있는 위험이 존재하기에 본 논문에서는, 촬영된 영상 내의 개인정보 및 민감정보를 최신 기술 기반 딥뉴럴

네트워크를 활용하여 포착하여 삭제하고 삭제된 부분은 인공지능 기술을 기반으로 물체가 없는 실제 공간의 영상으로 교체하는 것이 연구의 목표입니다. 1)한국형 민감물체 정의, 2)민감 물체 Detection 및 segmentation 과 inpainting 구현, 3) 한국형 부동산환경에 맞는 domain adaptation 성능 향상 및 dataset 유형에 따른 비용감소.

## 2. Method

한국 주거지 특성을 조사 및 분석하여, 한국 고유의 부동산 내 민감 물체를 연구하였다. 단일 RGB 영상을 사용하여, 최신 딥 뉴럴넷 알고리즘 기반 고정밀 민감 객체 검출 및 knowledge-based structure-preserving semantic scene reconstruction 를 연구개발 하였다.

이를 위해, 자체 학습 database 구축하고, YOLO (You Only Look Once) v5 및 LAMA (Large Mask inpainting) 알고리즘을 활용하여 최종 privacy preserving proptech 기술 완성하였다.

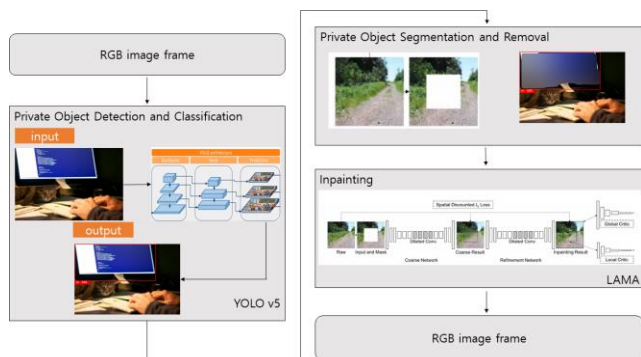


그림 1. Proposed algorithm overflow

### 2.1 한국형 Privacy Object DB 정의 및 구축

Untact (비대면)시대에 한국형 부동산의 Sensitive Object 시장조사를 통해, “TV, 컴퓨터, 태블릿, laptop 화면”을 민감 object 대표 물체로 선정하였으며 학습 Dataset 구축을 위해 Google 에서 Monitor 사진 및 Monitor 가 포함된 office 사진 약 1000 장을 Web Crawling 을 통해 받아 각각의 image labeling 을 하였다. MS COCO DB 로 학습된 pretrained model 적용시 한국 가정집/오피스 환경과 domain gap 차이로 인한 성능 저하가 발생하여 한국형 부동산 환경에 맞는 인공지능 모델을 구축하였다.

### 2.2 Privacy Object Detection

RGB 영상으로부터 물체 및 사람을 검출하는 기술인 Object detection 을 통해 고정밀 객체 검출 및 객체의 3D pose 추정을 하였다. 민감 object(monitor)에 영상 DB 를 활용하고 최신

딥러닝 뉴럴넷 기반 object detection 알고리즘 YOLOv5 [3]를 활용하였다.

YOLO(You Look Only Once)란 이미지에서 특정 object 의 영역을 찾아 bounding box regression 을 동시에 진행시키는 알고리즘으로 pretrained 된 head 를 통해서 feature map 을 얻고 back 에서 detection 을 한 후 classification 과 box regression 을 동시에 진행한다.

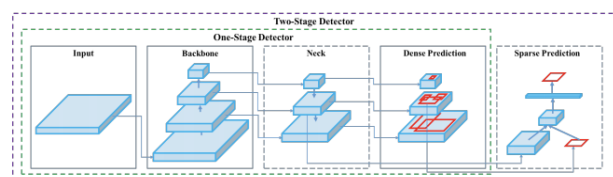


그림 2. YOLO v5 network architecture [3]

특정객체를 Box 로 위치 값을 알려주는 localization 알고리즘은 BBR(Bounding Box Regression)을 활용하며 각각의 Object 의 ground truth box 와 predict box 비교하여 mAP@0.6 만 사용하고 grid 를 통해 box 내부의 위치를 확인하여 정확한 confidence box 를 사용하여 classification 을 구할 수 있다. 기본적인 YOLO 알고리즘을 통해 기계학습을 진행하였다. Transfer learning for target domain adaptation 을 위하여 pretrained model 에 사용된 DB 의 feature distribution 수정을 위한 다양한 추가 DB 학습을 실험을 통해 모델 최적화를 진행하였다.

### 2.3 Object Inpainting with Segmentation

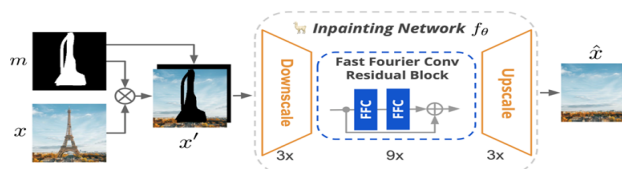


그림 3. Lama 의 frame work [4]

Detect 된 민감 object 에 대해, 영상 내의 민감 object 를 단순 삭제할 시에는, 부자연스러운 영상 생성으로 인해 물체 및 사람에 대한 의미론적으로 만족스럽지 못한 영상 결과가 나온다. 상용화를 위해서는, 최대한 자연스러운 부동산 매물 영상 제공을 위해, 원본 영상의 scene layout 을 보존해주는, 인공지능 기반 structure-preserving inpainting 이 필요하다. 본 연구에서는, 최신 뉴럴넷 기반 LaMa Image Inpainting (Resolution-robust Large Mask Inpainting with Fourier Convolutions) [4] 알고리즘을 활용하여, 사생활이 노출되는 모니터 디스플레이 영상에서 자연스럽게 디스플레이 안의 정보를 자연스럽게 메꿔주는 알고리즘을 연구 개발하였다.

본 연구에서 사용한 LaMa Inpainting 은 [4] FFC(fast Fourier convolution) block 을 이용하여 기존의 encoder, decoder 방식을 대신한다. 이는 기존의 Image Inpainting 에서는 large receptive field 가 적용이 잘 되지 않았던 점을 해결하여 보다 넓은 마스크를 가진 영상에서 타 모델들에 비해 뛰어난 성능을 보여준다. 그리고 adversarial loss[5]와 high receptive field perceptual loss[6]를 결합한 multi-component loss를 제시하여 성능 향상을 이끄는 feed-forward ResNetlike inpainting network inpainting 알고리즘이다.

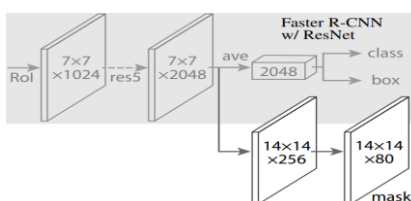


그림 4. Mask R-CNN 의 frame work [7]

Inpainting 알고리즘의 성능을 개선하기 위하여, image segmentation module 을 추구하고자 자동으로 privacy object 의 mask 생성하여 Inpainting 알고리즘에 적용하여 전체 성능을 개선하였다. 본 연구에서 영상 속 민감정보를 Segmentation 을 하기 위한 방법으로 Mask R-CNN [7]에서 segmentation module 추출하여 본 YOLO v5 및 LAM 알고리즘에 통합하여 전체 software 를 python 으로 구현하였다. Mask R-CNN 은 기존 object detection 모델인 Faster R-CNN 에서 object mask branch 가 추가된 모양을 하고 있다. 이전의 다른 Instance segmentation 모델들처럼 예측한 mask 에 따라 class 를 나누지 않고, Faster R-CNN 과 같이 class 와 bounding box 의 예측 과정과 binary mask 의 예측 과정을 병렬로 연결하여 더 나은 성능을 보여준다 거기에 더해 detect 한 물체의 mask 를 훨씬 더 세밀하게 추출해 내기 위하여 Faster R-CNN 에서 사용하던 RoI-Pooling 대신 새로운 layer 를 사용한다. Mask R-CNN 에서는 quantization loss 를 줄여 정확한 masking 을 출력하기 위해 RoI-Align layer 를 제안하였는데, RoI-Align 은 feature map 에서 RoI 의 경계부분에 해당하는 bin 의 feature 값을 bilinear interpolation 을 사용하여 정확한 값을 계산한 후 pooling 을 진행하여 최종 출력되는 binary mask 의 위치가 실제 RoI 와 거의 일치하도록 한다.

### 2.4 Transfer Learning- Domain Adaptation

Public Dataset (MS COCO DB)으로 학습된 pretrained model 적용시, 한국 가정집/오피스 환경과 domain gap 차이로 인한 성능 저하가 발생하였다. 이를 해결 하기위해 한국형 부동산

환경(영상 촬영 조도, 디스플레이 스타일, 주변 가구 배치 등)에 맞는 인공지능 모델 구축을 필요로 하여 Transfer learning for target domain adaptation 을 사용하였고 이를 위하여 pretrained model 에 사용된 DB 의 feature distribution 수정을 위한 다양한 추가 DB 학습을 실험하였다.

표 1 Transfer learning for target domain adaption 을 위하여 pretrained model 에 사용된 DB 의 feature distribution 수정 위한 다양한 추가 DB 학습 실험

Training DB	# of dataset
MS COCO (범용 public)	7,000
Web Crawling (한국형 부동산)	1,500
자체 구축 DB (한국형 부동산)	1,000

### 3. Experimental Results

민감 object detection 및 inpainting algorithm 을 적용하여 자연스러운 privacy preserving PropTech 기술을 연구 개발하였다. 본 알고리즘은 Google Colab [ref] GPU (Tesla T4, 15110MiB) 환경에서 1920 x 1080 입력영상으로부터 YOLO v5 privacy object detection 및 LAMA inpainting 을 파이썬 언어와 Pytorch framework 으로 구현하였다. 다양한 YOLO v5 를 정밀도 및 속도 test 실험하였고 최종적으로 YOLO v5 medium 버전으로 구현하였다. Proposed algorithm 은 일반적인 환경에서 98%의 정확도와 challenge 한 환경에서(occlusion 빛 반사, 저조도 등) 81% 정확도를 보였다 (표 2 및 그림 5, 6). 학습 dataset 구축을 위하여 Google 에서 한국 모니터 사진 및 모니터가 포함된 office 사진 약 1500 장을 모으고 1000 장은 자체 촬영하여 Labeling 등을 거쳐 database 를 구축하였다.

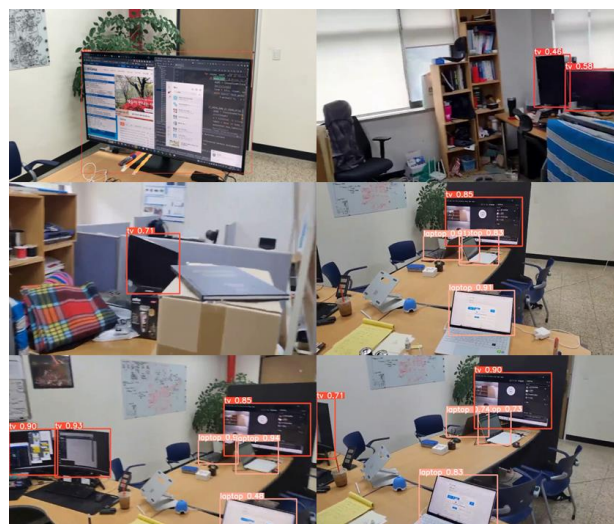


그림 5. Privacy Object detection result (디스플레이)

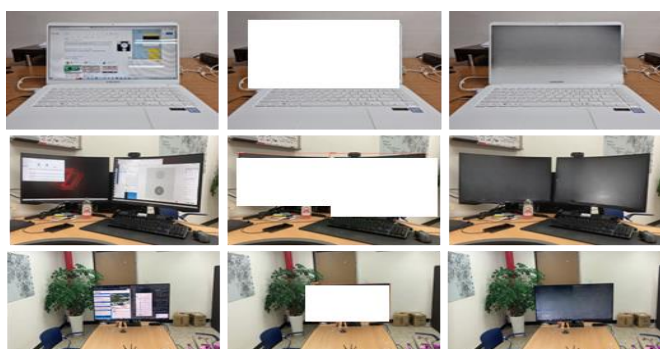


그림 6. Privacy Preserving results (디스플레이 정보 삭제 및 복원), Input(좌) Privacy Removed(중), Proposed Inpainted result(우)

표 2 Proposed algorithm accuracy results

	Pretrained Model	Customed Model 1	Customed Model 2
Test #1 (MS COCO DB 337장)	86%	87%	<b>89%</b>
Test #2 (한국 부동산 DB 1000장) - 일반적 환경	94%	96%	<b>98%</b>
Test #3 (한국 부동산 DB 1000장) - Challenge 환경	77%	75%	<b>81%</b>

기존의 MS COCO DB 와 Custom DB 를 비교하기 위해서 Precision 과 Recall 을 사용하였다. Precision 은 객체를 검출한 것 중에 올바르게 검출된 개수를 나타내는 것이고, Recall 은 검출해내야 하는 정답 객체중에 객체라고 올바르게 검출된 개수를 나타낸 것이다. (그림 7)

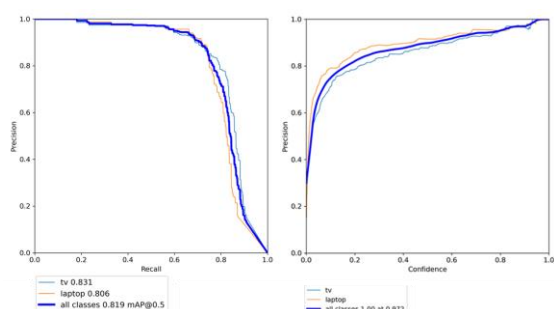


그림 7. Proposed algorithm training precision-recall curves (left) and precision-confidence curves (right)

#### 4. Discussion and Conclusion

본 privacy-preserving 연구방법론은, 사용자에게 자연스러운 결과 영상을 제공은 inpainting 알고리즘의 성능에 의존하다. 본 연구에서 활용한 LAMA-based inpainting 알고리즘은 지우고자 하는 mask 이미지의 결과에 따라 자연스러운 영상 reconstruction 성능의 차이가 발생한다 (그림 8). 본 연구 결과의 성능 향상을 위해서는, 고정밀 image segmentation mask generation 연구 혹은 rough mask 로도 고정밀 성능을 낼 수 있는 inpainting 알고리즘에 대한 연구가

필요하며, 추후 추가 연구를 진행하여 해결하고자 한다.

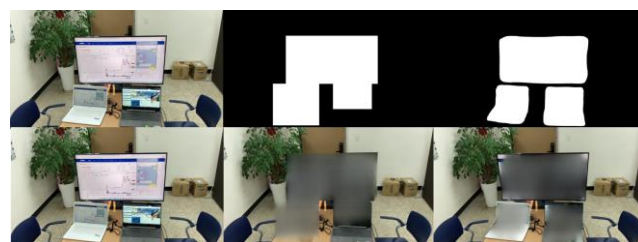


그림 8. Algorithm performance comparison: 1st row: original image (left), bbox area (middle), segmented mask (right). 2nd row: original image (left), inpainted image with bbox from YOLO v5 (middle), inpainted image with segmented mask from Mask R-CNN (right)

메타버스 부동산 중개 플랫폼 내 Untact(비대면) 한국형 부동산의 개인정보 보호를 위한 부동산 민감 물체 검출, 삭제 및 복원해주는 privacy-preserving Proptech(부동산) 기술에 대해 연구 개발을 완료하였다. 최신 뉴럴넷 인공지능 네트워크 기반인 YOLO v5, LAMA 기술을 활용하여, 개인 display 정보에 대해 민감 정보 검출, 삭제 및 복원에서 일반적인 환경에서 98%, challenge 환경에서 81%의 높은 성공률을 보여주며 본 논문의 연구를 수행하였다.

#### References

- [1] Turchetti, Lorenzo. "The impact of technological advances on the real estate industry: a focus on proptech and innovative business models." (2020).
- [2] Boeing, Geoff, et al. "Housing search in the age of big data: smarter cities or the same old blind spots?." Housing Policy Debate 31.1 (2021): 112-126.
- [3] Glenn Jocher et al. yolov5. <https://github.com/ultralytics/yolov5>, 2021. 1, 2, 3, 5, 6
- [4] Suvorov, Roman, et al. "Resolution-robust large mask inpainting with fourier convolutions. " Proceedings of the IEEE/CVF Winter Conference on Application s of Computer Vision, 2022.
- [5] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision, pages 694-711. Springer, 2016.
- [6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, 2017.
- [7] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.