사물인터넷융복합논문지 Vol. 11, No. 6, pp. 175-180, 2025 https://doi.org/10.20465/KIOTS.2025.11.6.175

CNN 기반 영상 분류 성능에 대한 전이학습 요인 분석

안우중[1], 나인섭[2*]
[1]전남대 데이터사이언스대학원 학생, [2]전남대 데이터사이언스대학원 교수

Analysis of Transfer Learning Factors in CNN-Based Image Classification Performance

Woo Jung Ahn[1], In Seop Na[2*]
[1]Student, Chonnam National University, Graduate School of Data Science
[2]Professor, Chonnam National University, Graduate School of Data Science

요 약 본 연구는 합성곱 신경망(CNN) 기반 전이학습에서 도메인 유사도, 백본 복잡도, 전이 전략이 영상 분류 성능에 미치는 영향을 실험적으로 분석하였다. CIFAR-10을 소스 도메인으로 사전학습한 후, STL-10(유사 도메인)과 Fashion-MNIST(비유사 도메인)에 대해 VGG16과 InceptionV3 백본을 사용하여 Feature Extraction과 Fine-Tuning 전략을 적용한 2×2×2 요인 설계 실험을 수행하였다. 실험 결과, 유사 도메인에서는 InceptionV3와 Fine-Tuning 조합이 정확도 90.5%로 최고 성능을 보였으며, 비유사 도메인에서는 VGG16과 Feature Extraction 조합이 78.6%의 안정적인 성능을 유지하였다. 본 연구는 전이학습의 성능이 단일 요인이 아닌 세 요인 간 상호작용에 의해 결정됨을 실증하였으며, 도메인 적응 및 소규모 데이터 전이학습 설계 시 실질적인 가이드라인을 제시한다.

주제어 : 전이학습, CNN, 도메인 유사도, Fine-Tuning, Feature Extraction, 영상 분류

Abstract This study experimentally analyzed the effects of domain similarity, backbone complexity, and transfer strategy on image classification performance in CNN-based transfer learning. After pre-training on CIFAR-10 as the source domain, a 2×2×2 factorial design experiment was conducted applying Feature Extraction and Fine-Tuning strategies using VGG16 and InceptionV3 backbones on STL-10 (similar domain) and Fashion-MNIST (dissimilar domain). The experimental results showed that the combination of InceptionV3 and Fine-Tuning achieved the highest performance with 90.5% accuracy in the similar domain, while the combination of VGG16 and Feature Extraction maintained stable performance of 78.6% in the dissimilar domain. This study demonstrated that transfer learning performance is determined by the interaction among three factors rather than a single factor, and provides practical guidelines for selecting strategies in domain adaptation and small-scale data transfer learning design.

Key Words : Transfer Learning, CNN, Domain Similarity, Fine-Tuning, Feature Extraction, Image Classification

*교신저자 : 나인섭(ypencil@hanmail.net)
접수일 2025년 11월 28일 수정일 2025년 12월 14일 심사완료일 2025년 12월 17일

www.kci.go.kr

1. 서론

딥러닝 기반 이미지 분류 기술은 의료 영상, 자율주행, 보안 감시 등 다양한 분야에서 활용되고 있다[1,2]. 특히 합성곱 신경망(CNN)을 중심으로 한 모델들은 대규모 데이터셋을 활용하여 높은 정확도를 달성하고 있으나, 막대한 양의 학습 데이터와 고성능 연산 자원에 의존한다는 한계가 있다[3,4].

이러한 문제를 해결하기 위한 방안으로 전이학습(Transfer Learning)이 주목받고 있다. 전이학습은 소스도메인에서 사전 학습된 지식을 타깃 도메인으로 전이함으로써 데이터 부족 문제를 완화하고 학습 효율을 향상시키는 방법이다[5,6]. 그러나 도메인 간 차이가 클 경우 도메인 불일치(domain shift) 현상이 발생하여 성능이 저하되는 문제가 있다[7,8].

기존 연구들은 전이학습의 성능 향상을 위해 공유 가중치 구조 설계, 정규화 기법 개선, 특징 보강 등 개별 요소 기술에 집중해왔다[9-11]. 그러나 도메인 유사도, 백본 복잡도, 전이 전략의 상호작용이 성능에 미치는 영향을 종합적으로 분석한 연구는 부족한 실정이다.

본 연구는 이러한 한계를 보완하기 위해 CIFAR-10을 소스 도메인으로, STL-10(유사 도메인)과 Fashion-MNIST(비유사 도메인)를 타깃으로 설정하였다. VGG16과 InceptionV3를 백본으로 사용하고, Feature Extraction과 Fine-Tuning 두 가지 전이 방식을 동일한 조건에서 비교·분석하여 전이학습 전략 선택의 기준을 제시하고자 한다.

2. 이론적 배경

2.1 전이학습과 전이 전략

전이학습은 한 도메인에서 학습된 지식을 다른 도메인으로 전이하여 새로운 문제 해결에 활용하는 학습 방식이다[5]. Pan과 Yang(2010)은 전이학습을 소스와 타깃의 관계에 따라 Inductive, Transductive, Unsupervised 형태로 구분하였다[12].

전이학습의 대표적 전략으로는 Feature Extraction과 Fine-Tuning이 있다. Feature Extraction은 사전학습된 백본의 합성곱 계층을 고정한 채 상위 분류기만 학습하는 방식으로, 연산 효율이 높고 소량 데이터에서도 안정적이다[13].

반면 Fine-Tuning은 사전학습된 모델의 일부 또는 전체 계층을 재학습하여 타깃 도메인의 통계적 특성에 적응시키는 방법이다[14]. 선행 연구에서는 도메인 간 차이가 클 경우 Fine-Tuning이 효과적일 수 있다고 제시되었으나, 이는 대규모 사전학습 데이터, 충분한 타깃 데이터, 강한 정규화 등의 전제 조건이 만족될 때 유효하다. 소규모 사전학습 환경이나 도메인 구조가 근본적으로 상이한 경우, Fine-Tuning은 오히려 과적합을 유발하여 성능 저하를 초래할 수 있다.

Yosinski et al.(2014)은 CNN의 저수준 계층은 에지, 텍스처 등 일반적 특징을 학습하여 도메인이 다르더라도 재사용 가능하지만, 고수준 계층은 도메인 특이적 표현을 학습하므로 세밀한 재적합이 필요하다고 밝혔다[1].

Kolesnikov et al.(2020)은 Big Transfer(BiT) 연구를 통해 초대규모 자연영상 데이터셋에 기반한 사전학습이 다양한 다운스트림 과제에서 강력한 전이 성능을 제공함을 보여주었다. 사전학습 규모가 전이 성능의 핵심 요소임을 실증적으로 제시한 연구로 평가된다[17]. Raghu et al.(2021)은 Vision Transformer(ViT)의 내부 표현 구조와 계층적 특성을 분석하여 CNN 기반 전이학습과의 근본적인 차이를 규명하였다. 이러한 결과는 백본 구조 선택이 단순한 정확도 비교를 넘어 도메인 특성과 데이터 규모에 따라 달라져야 함을 시사한다[18].

2.2 도메인 유사도와 백본 복잡도

도메인 유사도는 두 데이터셋이 시각적·통계적 특성에서 얼마나 유사한지를 나타내는 척도이며, 유사도가 높을수록 소스에서 학습된 표현이 타깃에서도 효과적으로 작동할 가능성이 커진다[15]. 반대로 도메인 쉬프트가 크면 기존 표현이 타깃의 통계 구조를 포착하지 못해 부정적 전이(negative transfer) 위험이 증가한다[7].

백본 구조의 복잡도도 전이 성능에 중요한 영향을 미친다. VGG16은 단순한 16층 구조로 안정적 전이 성능을 보이며, InceptionV3는 다중 스케일 특징을 병렬적으로 학습하는 복합적 구조로 고차원적 시각 표현에 강점을 가진다[2,16].

3. 연구방법

3.1 실험설계

본 연구는 2×2×2 요인 설계로 구성되었다. 첫 번째 요인은 도메인 유사도(STL-10, Fashion-MNIST), 두 번째 요인은 백본 복잡도(VGG16, InceptionV3), 세 번째

www.kci.go.kr

요인은 전이 전략(Feature Extraction, Fine-Tuning)으로, 총 8개 조합을 실험하였다.

〈Table 1〉 Summary of Experimental Design

Factor	Level	Description
Domain Similarity	2	STL-10(Similar), Fashion-MNIST(Dissimilar)
Backbone Complexity	2	VGG16, InceptionV3
Transfer Strategy	2	Feature Extraction, Fine-Tuning
Total Number of Experiments	8	2×2×2 Factorial Design

3.2 데이터셋 구성

본 연구에서는 CIFAR-10, STL-10, Fashion-MNIST의 공식 제공된 학습·검증·테스트 분할을 그대로 사용하였다. CIFAR-10은 학습용 50,000장과 테스트용 10,000장으로 구성되며, STL-10은 학습용 5,000장과 테스트용 8,000장, Fashion-MNIST는 학습용 60,000장과 테스트용 10,000장으로 이루어져 있다.

〈Table 2〉 Comparison of Dataset Composition

Type	Source	Target 1	Target 2
Dataset	CIFAR-10	STL-10	Fashion-MNIST
Resolution	32×32	96×96	28×28
Channels	3	3	1 -> 3
Number of Classes	10	10	10
Training Samples	50,000	5,000	60,000
Test Samples	10,000	8,000	10,000
Domain Similarity	Criterion	High	Low

각 데이터셋은 원본 제공 시점에서 이미 학습(Training)과 테스트(Test) 데이터가 분리되어 있으며, 본 연구에서는 이러한 공식 분할을 변경하지 않고 그대로 사용하였다. 따라서 학습 데이터와 테스트 데이터는 중복 없이 완전히 독립적으로 유지되며, 어떠한 이미지도 두 집합에 동시에 포함되지 않는다. 이는 모델 평가의 공정성과 재현성을 확보하기 위함이다.

CIFAR-10을 소스 도메인으로 설정하고, STL-10(유사 도메인)과 Fashion-MNIST(비유사 도메인)를 타깃 도메인으로 사용하였다. CIFAR-10과 STL-10은 모두 자연 이미지로 배경·색·형태 특성이 유사하며, Fashion-MNIST는 단색 대비·윤곽 중심의 의류 이미지로 도메인 차이가 크다.

모든 데이터셋은 백본 입력 규격에 맞춰 224×224 픽셀로 리사이즈하였으며, Fashion-MNIST는 1채널 이미지를 3채널로 복제하였다.

3.3 실험환경 및 모델 학습 설정

본 연구의 모든 실험은 Google Colab Pro+ 환경에서 수행되었으며, 학습에는 NVIDIA T4 GPU(16GB VRAM)를 사용하였다. T4 GPU는 딥러닝 학습에 널리 사용되는 GPU로, 합리적인 FP32 연산 성능과 충분한 메모리를 제공하여 VGG16 및 InceptionV3와 같은 중·대형 CNN 모델 학습에 적합하다. 또한 2.30GHz의 Intel Xeon Cascade Lake 기반 vCPU와 27GB 시스템 메모리를 바탕으로 데이터셋 로딩, 전처리 및 배치 구성 과정에서 병목현상 없이 안정적인 수행이 가능하다. 딥러닝 프레임워크는 Python 3.10 환경에서 PyTorch 2.1.0과 TorchVision 0.16 버전을 사용하였다.

Feature Extraction 단계에서는 Adam Optimizer (learning rate=1e-3)를 사용하였고, Fine-Tuning 단계에서는 AdamW(learning rate=5e-5~1e-4)를 적용하였다. 학습률 스케줄링에는 Cosine Decay 기법을 사용하였으며, 5 Epoch의 Warm-up 단계를 추가하였다.

정규화는 L2 정규화(Weight Decay=1e-4), Dropout(0.3 to 0.5), Label Smoothing(0.05)을 병행하였다. 학습은 Batch Size=64 to 128, Epoch=50 to 100으로 진행되었으며, 검증 손실이 10 Epoch 연속 개선되지 않을 경우 Early Stopping을 적용하였다.

3.4 평가지표

모델 성능은 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-Score로 평가하였다. 또한 혼동행렬(Confusion Matrix)을 통해 클래스별 오분류 경향을 분석하였다.

4. 실험 결과

4.1 사전학습 결과

〈Table 3〉은 CIFAR-10 데이터셋을 이용한 사전학습 결과를 요약한 것이다. 두 백본 모델의 평균 정확도는 0.92, F1-Score는 0.91로 나타나 전이학습의 안정적인 기준선을 확보하였다.

www.kci.go.kr

〈Table 3〉 Summary of Pretraining Results on CIFAR-10

Indicator	Average
Accuracy	0.92
Precision	0.91
Recall	0.91
F1-Score	0.91

4.2 전이학습 성능 비교

〈Table 4〉는 8개 실험 조합의 성능을 요약한 것이다. 유사 도메인(STL-10)에서는 Fine-Tuning 전략이, 비유사 도메인(Fashion-MNIST)에서는 Feature Extraction 전략이 우수한 성능을 보였다.

〈Table 4〉 Comparison of Transfer Learning Performance

Domain	Backbone	Strategy	Accuracy	Precision	Recall	F1-score
STL-10 (유사)	VGG16	FE	0.832	0.824	0.801	0.812
	VGG16	FT	0.861	0.856	0.829	0.841
	InceptionV3	FT	0.883	0.872	0.862	0.867
	InceptionV3	FE	0.905	0.893	0.876	0.884
Fashion-MNIST (비유사)	VGG16	FE	0.786	0.783	0.760	0.771
	VGG16	FT	0.758	0.757	0.741	0.749
	InceptionV3	FE	0.749	0.751	0.746	0.748
	InceptionV3	FT	0.724	0.739	0.725	0.732

4.2.1 전이 전략별 성능

Fine-Tuning 전략은 유사 도메인에서 평균 정확도가 Feature Extraction 대비 2.2%p 높았으며, 특히 InceptionV3와 결합 시 90.5%로 최고 성능을 기록하였다. 이는 사전학습된 가중치를 미세 조정하여 소스와 타깃의 시각적 특징이 일치할 때 긍정적 전이가 발생함을 보여준다.

반면 비유사 도메인에서는 Feature Extraction이 더 안정적이었다. VGG16-Feature Extraction은 78.6%의 정확도를 유지한 반면, InceptionV3-Fine-Tuning은 72.4%로 하락하여 복잡한 구조가 과적합을 유발할 수 있음을 나타냈다.

4.2.2 도메인 유사도별 성능

유사 도메인(STL-10)에서는 모든 조합이 비유사 도메인보다 높은 성능을 보였다. 이는 소스(CIFAR-10)와의 시각적 유사성이 높을수록 전이된 특징 표현이 효율적으로 작동함을 의미한다.

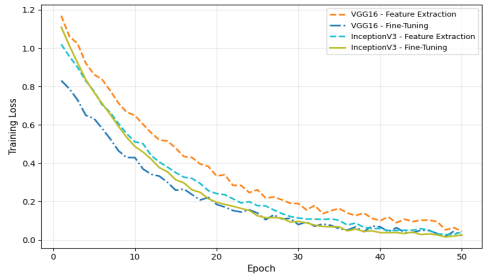
비유사 도메인(Fashion-MNIST)에서는 단순 백본(VGG16)과 Feature Extraction 조합이 가장 안정적이었다. 이는 저복잡도의 네트워크가 과적합 위험을 줄이고, 도메인 간 차이가 큰 환경에서도 일반화 성능을 유지할 수 있기 때문이다.

4.2.3 백본별 성능

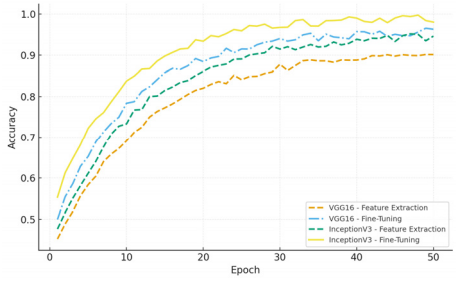
InceptionV3는 유사 도메인에서 뛰어난 성능을 보였으나, 비유사 도메인에서는 과적합 경향을 보였다. VGG16은 두 도메인 모두에서 고른 성능을 유지하며, 특히 비유사 도메인에서 안정적이었다.

4.3 학습과정 분석

Fig.1 학습 손실과 Fig .2 정확도 변화 추이 분석 결과, Fine-Tuning 전략은 초기 학습 속도와 수렴 속도 모두에서 Feature Extraction보다 우수하였으나, 도메인 차이가 클 경우 오히려 성능 저하가 발생하였다.



[Fig. 1] Training Loss Curves for Each Model

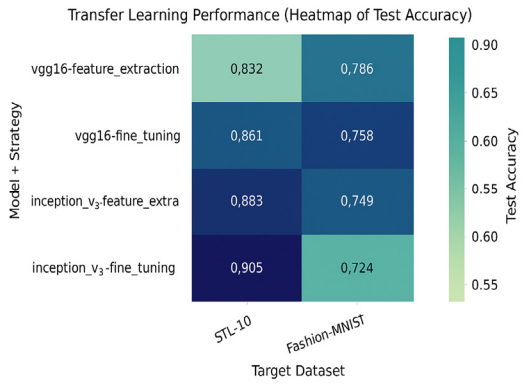


[Fig. 2] Accuracy curves for each model

4.4 혼동행렬 분석

STL-10에서는 InceptionV3-Fine Tuning이 0.905의 정확도로 가장 우수한 전이 성능을 보였다.

이는 고복잡도 백본과 미세 조정 전략이 유사 도메인에 효과적으로 적응했음을 의미한다.



[Fig. 3] Transfer Learning Performance (Heatmap Of Test Accuracy)

반면 VGG16-Feature Extraction은 0.832로 다소 낮지만 안정적인 성능을 유지하였다.

Fashion-MNIST에서는 전체적으로 정확도가 하락했으며 InceptionV3-Fine Tuning은 0.724로 성능 저하가 나타났다. 이는 도메인 차이가 클수록 복잡한 모델이 일반화에 어려움을 겪음을 보여준다.

반대로 VGG16-Feature Extraction은 0.786으로 비유사 도메인에서도 가장 높은 성능을 기록하였다.

5. 결론 및 향후 연구

5.1 결론

연구는 CNN 기반 전이학습에서 도메인 유사도, 백본 복잡도, 전이 전략의 세 요인이 영상 분류 성능에 미치는 영향을 실험적으로 분석하였다. 주요 연구 결과는 다음과 같다.

첫째, 도메인 유사도가 높을수록 Fine-Tuning 전략이 유의미한 성능 향상을 보였으며, InceptionV3-Fine-Tuning 조합은 유사 도메인에서 정확도 90.5%로 최고 성능을 기록하였다.

둘째, 비유사 도메인에서는 Feature Extraction 전략이 더 안정적이었으며, VGG16-Feature Extraction 조합이 78.6%의 높은 일반화 성능을 유지하였다.

셋째, 전이학습의 성능은 단일 요인이 아닌 세 요인 간 상호작용에 의해 결정되며, 유사 도메인에서는 복잡한 백본과 Fine-Tuning이, 비유사 도메인에서는 단순 백본과 Feature Extraction이 최적의 조합임을 실증하였다.

〈Table 5〉 Optimal Combination of Domain and Backbone

Domain	Similar Domain (STL-10)	Dissimilar Domain (Fashion-MNIST)
Optimal Combination	InceptionV3 + Fine-Tuning	VGG16 + Feature Extraction
Performance Summary	High Accuracy (90.5%) & Deep Representational Layers	Stable Performance(78.6%), Overfitting Minimization

본 연구의 결과는 향후 도메인 적응 및 소규모 데이터 전이학습 설계 시 전략 선택에 실질적인 가이드라인을 제공할 수 있다.

5.2 향후 연구

향후 연구에서는 다음과 같은 확장이 필요하다. 첫째, 의료영상, 위성영상 등 도메인 간 차이가 극단적인 환경에서의 전이학습 효율성 검증이 필요하다. 둘째, ResNet, EfficientNet, Vision Transformer 등 최신 백본 구조와의 비교 연구를 수행할 필요가 있다. 셋째, 3요인 혼합 분산분석(Three-way ANOVA)을 통한 통계적 유의성 검정으로 연구 신뢰도를 향상시킬 수 있다. 넷째, FID와 MMD 등 정량적 도메인 유사도 지표를 적용하여 전이 성능 예측 모델을 개발할 수 있다.

REFERENCES

[1] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? Advances in Neural Information Processing Systems, 27.

[2] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818-2826.

[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097-1105.

[4] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.

[5] Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.

www.kci.go.kr

[6] Bengio, Y. (2012). Deep learning of representations for unsupervised and transfer learning. Proceedings of ICML Workshop on Unsupervised and Transfer Learning, 17-36.

[7] Ben-David, S., Blitzer, J., Crammer, K., Pereira, F. (2007). Analysis of representations for domain adaptation. Advances in Neural Information Processing Systems.

[8] Kornblith, S., Shlens, J., & Le, Q. V. (2019). Do better ImageNet models transfer better? Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2661-2671.

[9] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014). DeCAF: A deep convolutional activation feature for generic visual recognition. International Conference on Machine Learning, 647-655.

[10] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

[12] Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.

[13] Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1717-1724.

[14] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? Advances in Neural Information Processing Systems, 27.

[15] Kornblith, S., Shlens, J., & Le, Q. V. (2019). Do better ImageNet models transfer better? Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2661-2671.

[16] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning, 6105-6114.

[17] Kolesnikov, A., Beyer, L., Zhai, X., Puigcerver, J., Yung, J., Gelly, S., & Houlsby, N. (2020). Big transfer (BiT): General visual representation learning. In Computer Vision–ECCV 2020, 491–507. Springer.

[18] Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C., & Dosovitskiy, A. (2021). Do Vision Transformers See Like Convolutional Neural Networks? Advances in Neural Information Processing Systems (NeurIPS), 34, 12116–12128.

안 우 중(Ahn-Woo Jung) [정회원]

- 2022년 2월 : 광주대학교 공과대학 AI소프트웨어학과
- 2024년 2월 : 전남대학교 데이터사이언스대학원 (석사 재학중)

〈관심분야〉
인공지능, 영상처리, 사물인터넷

나 인 섭(Na-In Sup) [정회원]

- 1997년 2월 : 전남대학교 전산학과 졸업(이학사)
- 1999년 2월 : 전남대학교 전산통계학과 졸업(이학석사)
- 2008년 2월 : 전남대학교 전산학과 졸업(이학박사)
- 2023년 ~ 현재 : 전남대학교 문화콘텐츠학부, 데이터사이언스대학원 교수

〈관심분야〉
인공지능, 시각지능, 영상처리,
패턴인식객체(검출,분할,인식, 추적, 이해)

www.kci.go.kr