

## DINOv2: Görsel Temsil Öğreniminde Yeni Bir Yaklaşım

### 1. Giriş

Görsel temsil öğrenimi, görüntü sınıflandırma, nesne tanıma, segmentasyon ve benzeri birçok bilgisayarla görme (computer vision) görevinde kritik rol oynar. Son yıllarda, denetimsiz öğrenme (self-supervised learning - SSL) yöntemleri, özellikle etiketlenmemiş büyük veri kümeleriyle daha etkili modeller geliştirmeye olanak tanımıştır. Bu doğrultuda, Meta AI Research tarafından geliştirilen **DINOv2**, yüksek kaliteli görsel temsiller üreten, ölçeklenebilir, çok yönlü ve denetimsiz bir görsel öğrenme yöntemidir.

### 2. DINOv2 Nedir?

DINOv2 (Distillation with NO labels version 2), **Vision Transformer (ViT)** tabanlı bir **self-supervised learning** mimarisidir. DINOv2, tıpkı ilk versiyonu olan DINO gibi, etiketli veri gerektirmeden güçlü temsiller öğrenebilir; ancak bu versiyon, daha büyük veri kümeleri, gelişmiş eğitim teknikleri ve daha iyi genel performans sağlayan model güncellemeleriyle önemli gelişmeler sunmaktadır.

DINOv2'nin en temel amacı, görsellerden güçlü, genelleştirilebilir ve transfer edilebilir özellikler (features) çıkartmaktır.

### 3. Mimari Özellikler

- **Temel Bileşen:** Vision Transformer (ViT) mimarisi kullanılır (özellikle ViT-Small, ViT-Base, ViT-Large ve ViT-Giant varyantları).
- **Distillation (Damıtma) Yaklaşımı:** Öğrenci (student) ve öğretmen (teacher) ağları arasında bilgi aktarımı gerçekleştirilir. Her iki ağ da aynı mimariye sahip olsa da öğretmen ağı, öğrencinin geçmiş ağırlıklarının EMA (Exponential Moving Average) versiyonudur.
- **Çoklu Görsel Görünüm (Multi-crop strategy):** Aynı görüntünün farklı kırılmış versiyonları ile öğrenci ve öğretmen ağı eğitilir.
- **Augmentasyon:** RandAugment, Color jittering, Gaussian blur ve diğer gelişmiş veri artırma teknikleri ile modelin daha genelleşici hale gelmesi sağlanır.

### 4. Veri Kümesi ve Ölçeklenebilirlik

DINOv2, 1.2 milyar görsel içeren özel bir veri kümesi üzerinde eğitilmiştir. Bu veri kümesi, çeşitli web kaynaklarından elde edilmiş **etiketsiz** verilerden oluşur. Bu sayede DINOv2, yalnızca ImageNet gibi sınırlı etiketli veri kümeleri yerine daha geniş ve çeşitliliği yüksek görsellerle eğitilebilmiştir.

### 5. Özellik Öğrenme ve Transfer Edilebilirlik

DINOv2 ile eğitilmiş görsel temsiller:

- Görüntü sınıflandırma,
- Semantic ve instance segmentation,
- Feature matching,
- Video anlayışı,
- Zero-shot transfer, gibi birçok görevde başarıyla kullanılabilir. Özellikle downstream task'lerde fine-tuning yapmadan doğrudan kullanılacak kalitede özellikler sunar.

### 6. DINOv2'nin DINO'dan Farkları

Özellik	DINO	DINOv2
Mimari	ViT-S, ViT-B	ViT-L, ViT-G
Eğitim Verisi	ImageNet	1.2B görsel (etiketsiz, geniş kaynaklı)
Performans	Başarılı ama sınırlı	Daha iyi genelleme, daha güçlü temsil
Kullanım Alanı	Görsel temsiller	Multi-task görevlerde yüksek başarı

## 7. DINOv2 Kullanım Alanları

- **Görüntü Anlamlandırma:** Yüksek çözünürlüklü temsiller ile segmentasyon ve sınıflandırma işlemleri.
- **Bilgisayarla Görme Model Tabanı:** Özellikle etiketli veri az olduğunda transfer learning için temel özellik çıkarıcı.
- **Kümelenmiş Görsel Arama:** Görsel benzerlik tespiti (ör. CLIP benzeri görsel-sözel eşleme).
- **Robotik:** Görsel veri ile çevresel farkındalık.
- **Otonom Sistemler:** Yol ve nesne tanıma modülleri.

## 8. Deneysel Sonuçlar

DINOv2, birçok benchmark veri setinde (ImageNet-1K, COCO, ADE20K vb.) hem denetimli hem de diğer self-supervised yöntemlerle karşılaştırıldığında oldukça rekabetçi sonuçlar vermektedir. Özellikle:

- Linear probing testlerinde yüksek başarı (%80+ top-1 accuracy ImageNet'te),
- Semantic segmentation gibi görevlerde fine-tuning'e gerek kalmadan kullanılabilir temsiller,
- Daha az örnekle daha yüksek doğruluk sağlayan few-shot öğrenme başarısı göstermektedir.

## 9. Sonuç ve Değerlendirme

DINOv2, günümüz self-supervised görsel öğrenim yaklaşımlarında önemli bir aşama kaydeden, yüksek kaliteli ve esnek bir yöntemdir. Etiketsiz büyük veri kümeleri üzerinde eğitilerek, geniş görev yelpazesine kolayca entegre edilebilen temsiller üretmektedir. Derin öğrenme temelli görüntü işleme projelerinde, özellikle sınırlı etiketli veri koşullarında, güçlü bir başlangıç noktası sunar.

Ek:

- **DINOv2: Learning Robust Visual Features without Supervision** (Meta AI, 2023)
- Meta AI Blog: <https://ai.facebook.com/blog/dino-v2-self-supervised-learning-computer-vision/>
- GitHub: <https://github.com/facebookresearch/dinov2>