

BENCHMARK · 30 EYLÜL 2026

Hangi model, *ne kadar doğru, ne kadar hızlı?*

İki ölçüm. Önce tanıma servisinde kullanılabilecek altı görüntü modelini aynı veriyle, servisin kendi karar kuralıyla karşılaştırdık. Sonra kasaya tepeden bakan bir kamerayla gerçek kullanım açısını ölçtük. Aşağıdaki her rakam bu ölçümlerden gelir; ölçmediklerimizi de ayrıca yazdık.

Tarih	30 Eylül 2026
Referans galeri	419 görsel · 42 ürün sınıfı
Test seti	340 görsel · 34 ürün sınıfı
Veri	Grocery Store Dataset (MIT) · market rafı fotoğrafları
Karar kuralı	Servisle aynı: güven $\geq$ %95 ise sepete yazar
Makine	Apple M5 dizüstü · torch 2.14.0

01 · MODEL SEÇİMİ · RAF FOTOĞRAFI

MODEL DINOv2-base · Tüm kare · serviste çalışan

%85,9

Doğru ürün ilk tahminde

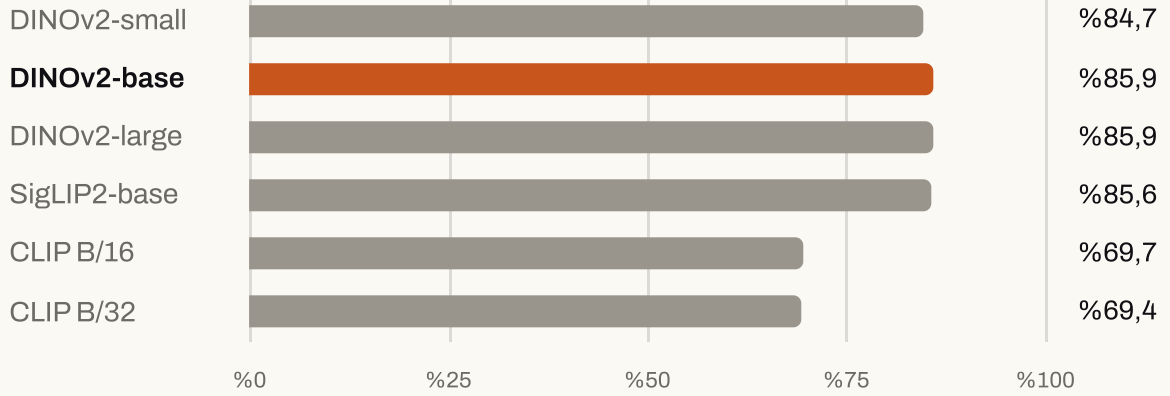
%97,4

Doğru ürün ilk 3 öneride

%38,8**%99,2**

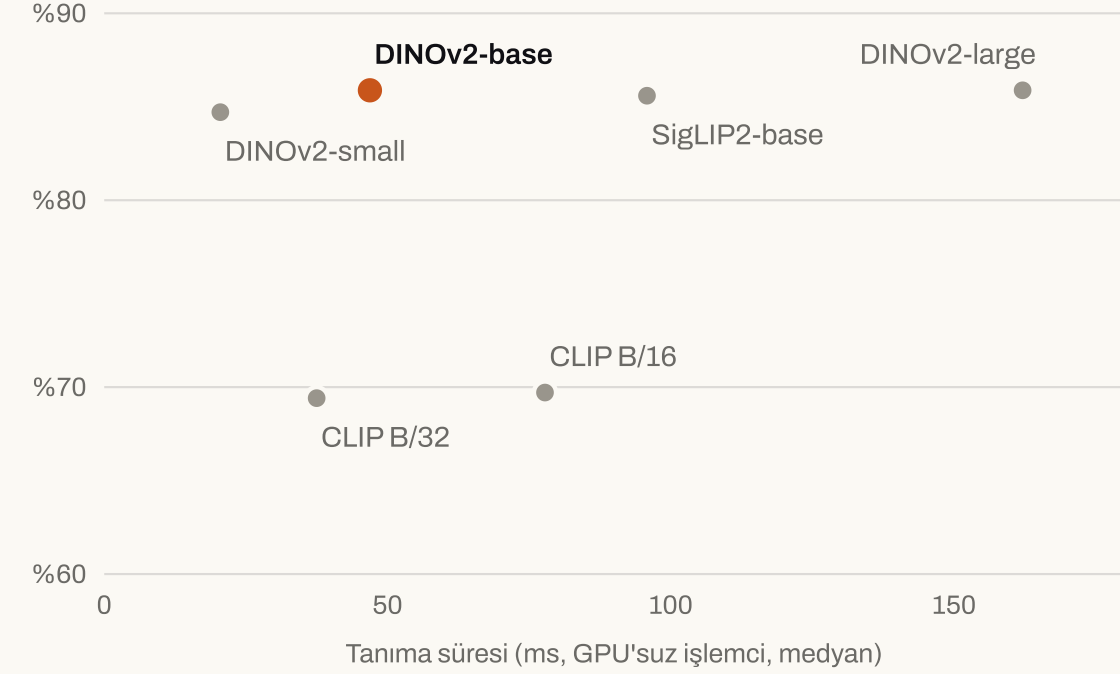
Modeller yan yana

Aynı test seti, aynı karar kuralı. Çubuk uzadıkça iyi.



Hız ve doğruluk

Sağa gittikçe yavaş, yukarı çıktıkça doğru. İdeal köşe sol üst.



İlk tahminde doğruluk · dikey eksen %60'tan başlar

Bir kalem kasada nasıl sonuçlanıyor?

Test setindeki 340 görselin dağılımı. DINOv2-base, tüm kare.



■ %38,8 Sormadan sepete yazıldı

■ %58,5 Kasiyere soruldu, doğru ürün adaylar arasındaydı

■ %2,6 Kasiyere soruldu, doğru ürün adaylarda yoktu

Sormadan yazılanlardan yanlış çıkan: 1 kalem.

Kırpım bu veride zarar veriyor

Görseldeki ürünü önce bir tespit modeliyle kırpıp sonra tanımak, raf fotoğraflarında doğruluğu düşürdü: tespit modeli yığından tek bir meyveyi kesiyor.

Tüm kare		%85,9
YOLO-World kırpımı		%76,5
OWLv2 kırpımı		%54,4
İlk tahmin · DINOv2-base		
Kutu bulduğu görsel oranı · YOLO-World		%35,7
Kutu bulduğu görsel oranı · OWLv2		%90,6

En çok karışan ürünler

Doğrusu → model ne dedi. Benzer görünen çiftler; servis bunlarda otomatik onay vermez, kasiyere sorar.

DOĞRUSU	MODEL	KAÇ KEZ
Golden elma	→ Yeşil elma	7
Nektarin	→ Şeftali	5
Limon	→ Misket limonu	4
Zencefil	→ Kırmızı pancar	3
Limon	→ Portakal	3
Mandalina	→ Portakal	3
Salatalık	→ Kabak	2
Mandalina	→ Limon	2

Süreler

Tek görsel, ısınmadan sonra, ön işleme dahil. Ağ ve kamera dahil değil.

ADIM	MODEL	İŞLEMCI P50	İŞLEMCI P95	APPLE GPU P50	APPLE GPU P95
Tanıma	DINOv2-small	20,5 ms	21,3 ms	9,9 ms	10,3 ms
Tanıma	DINOv2-base	46,9 ms	47,5 ms	22,5 ms	22,9 ms
Tanıma	DINOv2-large	162,1 ms	231,3 ms	68,9 ms	72,4 ms
Tanıma	SigLIP2-base	95,8 ms	126,4 ms	18,3 ms	18,5 ms
Tanıma	CLIP B/16	77,8 ms	101,3 ms	19,1 ms	20 ms
Tanıma	CLIP B/32	37,5 ms	44,1 ms	9,3 ms	9,9 ms
Tespit (kırpım)	YOLO-World	115 ms	168,9 ms	13,4 ms	15,4 ms
Tespit (kırpım)	OWLv2	1392,1 ms	1726,5 ms	653,1 ms	680,9 ms

Tüm sonuçlar

Model	Kırpım	İlk tahmin	İlk 3	Sormadan yazılan	Doğruluğu	Soruldu
DINOv2-small	Tüm kare	%84,7	%97,7	%33,2	%99,1	
DINOv2-base	Tüm kare	%85,9	%97,4	%38,8	%99,2	
DINOv2-large	Tüm kare	%85,9	%97,1	%46,2	%99,4	
SigLIP2-base	Tüm kare	%85,6	%99,7	%42,4	%100,0	
CLIP B/16	Tüm kare	%69,7	%95,3	%15,3	%100,0	
CLIP B/32	Tüm kare	%69,4	%92,3	%12,9	%97,7	
DINOv2-small	YOLO-World kırpımı	%75,3	%90,0	%19,7	%100,0	
DINOv2-base	YOLO-World kırpımı	%76,5	%91,8	%22,9	%98,7	
DINOv2-large	YOLO-World kırpımı	%79,4	%92,3	%35,6	%99,2	
SigLIP2-base	YOLO-World kırpımı	%72,1	%88,8	%23,8	%100,0	
CLIP B/16	YOLO-World kırpımı	%59,1	%84,4	%7,1	%100,0	
CLIP B/32	YOLO-World kırpımı	%55,0	%77,6	%6,5	%100,0	
DINOv2-small	OWLv2 kırpımı	%50,9	%65,3	%11,2	%97,4	
DINOv2-base	OWLv2 kırpımı	%54,4	%70,6	%16,2	%96,4	
DINOv2-large	OWLv2 kırpımı	%57,1	%70,9	%16,2	%96,4	
SigLIP2-base	OWLv2 kırpımı	%42,4	%59,1	%10,9	%97,3	
CLIP B/16	OWLv2 kırpımı	%36,2	%52,9	%2,4	%100,0	
CLIP B/32	OWLv2 kırpımı	%35,0	%48,5	%3,2	%100,0	

Kasada, tepeden *bakan kamerayla.*

Tezgâha tepeden sabitlenmiş bir kamerayla gerçek kullanımı taklit ettik. Referans galeri yine raf fotoğrafı; yani ilk satırlar, işletmenin hiç fotoğraf çekmediği kurulum gününü gösterir.

Tarih	2 Ekim 2026
Tekli ürün	40 kare · 11 ürün
Karışık tepsi	30 kare · 88 ürün
Referans galeri	419 raf fotoğrafı · 42 ürün sınıfı
İşletme fotoğrafı	Ürün başına en fazla 3; test edilen kare galeride hiç yok
Kamera	Telefon, sabit, tepeden · tek tezgâh

MODEL

%100,0

Kurulum günü: doğru
ürün ilk 3 öneride

%100,0

Ürün başına 3
fotoğraftan sonra: ilk
tahminde doğru

%93,3

Karışık tepside adet
doğru

Kurulum günü ve 3 fotoğraftan sonra

Terazide tek ürün. Servis yalnızca işletmenin kataloğunda arar; ortadaki satır bugünkü davranıştır.

Kurulum günü · tüm sınıflarda arama

İlk tahmin		%45,0
İlk 3 öneri		%85,0

Kurulum günü · işletme kataloğunda arama

İlk tahmin		%80,0
İlk 3 öneri		%100,0

Ürün başına 3 fotoğraftan sonra

İlk tahmin		%100,0
İlk 3 öneri		%100,0

Karışık tepside sayma

Aynı karede birden fazla ürün: önce ürünler kutulanır, her kutu ayrı tanınır. Ürün başına 3 fotoğraftan sonra, 30 tepsi, 88 ürün.

YOLO-World · ürün adlarıyla		%93,3
YOLO-World · genel		%46,7
OWLv2 · ürün adlarıyla		%53,3
OWLv2 · genel		%30,0
Boş kare farkı · lekeler		%16,7
Boş kare farkı · tek kutu		%0,0
Tüm kare · tek ürün		%0,0

Kurulum günü neyi neyle karıřtırdı?

Tüm sınıflarda arama. Çoğu, galerideki görselle kasadaki ürünün farklı görüldüğü makul karıřmalar (galerideki mandalinalar turuncu, kasadakiler yeřildi); iřletme kataloğunda aramak ya da 3 fotoğraf eklemek bunları çözüyor.

DOĞRUSU	MODEL	KAÇ KEZ
Yeřil elma	→ Golden elma	4
Kırmızı elma	→ řeftali	4
Mandalina	→ Misket limonu	4
Kırmızı elma	→ Golden elma	2
Salatalık	→ Kabak	2
Kırmızı elma	→ Nektarin	1

DINOv2-base · Tüm kare

Bu ölçümün söylemedikleri

01 · MODEL SEÇİMİ · RAF FOTOĞRAFI

- 01'deki görseller market rafından; kasaya tepeden bakan kamerayla yapılan ölçüm 02'de.
- 340 görselde bir puanlık fark yaklaşık 3 görsele denk gelir; ilk dört modelin ilk tahmin doğruluğu gürültü payı içinde eşit sayılmalı.
- Sunucu GPU'sunda süre ölçülmedi. Süreler bir dizüstü bilgisayarın işlemcisinden.
- Aynı karede birden fazla ürünü ayırıp sayma 01'in dışında; 02'de ölçüldü.
- Vektör veritabanı karşılaştırması (Qdrant, pgvector) yapılmadı; arama burada kesin hesapla yapıldı.

02 · KASA AÇISI

- Küçük bir set: 70 kare, tek tezgâh, tek ışık. Ürün bazında oranlar güvenilir değil; yön gösteren ilk sonuç olarak okunmalı.
- 3 fotoğraf testinde işletme fotoğrafları ve test kareleri aynı ürünlerden, aynı çekimden; sonuç iyimser.
- Katalogda arama sonuçlarında katalog 12 ürünlüktü; gerçek bir manav kataloğu daha büyük olur ve benzer ürünleri birlikte içerir.
- Kareler farklı videolardan alındı; kamera hafifçe kaydığı için boş kare farkı yöntemleri bu sette dezavantajlı.
- Pastane ürünleri bu ölçümde yok.

Kendiniz tekrar üretin

Docker veya vektör veritabanı gerekmez.

```
cd ai
# 01
python -m scripts.benchmark_models \
  --latency-devices cpu mps
# 02 · data/overhead
python -m scripts.benchmark_overhead
```

[docs/benchmarks/2026-09-30/](#) · [docs/benchmarks/2026-10-02-overhead/](#)