

## Özet

Bu makalede, tek bir görüntüye ilişkin Yüz Hizalama sorunu ele alınmaktadır. Yüzdeki referans noktaların (landmark) konumlarının, doğrudan piksel yoğunluğunun seyrek bir alt kümesinden tahmin etmek için bir regresyon ağacı dizisinin nasıl kullanılabilceği ve yüksek kaliteli tahminlere süper gerçek zamanlı performansla nasıl ulaşılabileceği gösterilmektedir. Karesel hata kayıplarının toplamını optimize eden ve veya doğal olarak eksik veya kısmen etiketli verileri işleyen bir regresyon ağacı dizisini öğrenmek için gradyan destekli genel bir çerçeve sunuyoruz. Görüntü verilerinin yapısını kullanan uygun öncelikleri (prior) kullanmanın, etkin detay (feature) seçimine nasıl yardımcı olduğunu gösteriyoruz. Farklı düzenleme stratejileri ve buna aşırı uyma ile mücadeledeki önemi de araştırılmıştır. Ayrıca, eğitim verisi miktarının öngörülerin doğruluğu üzerindeki etkisi analiz edilmiş ve sentezlenmiş veri kullanılarak veri sayısındaki artışın etkisini araştırılmıştır.

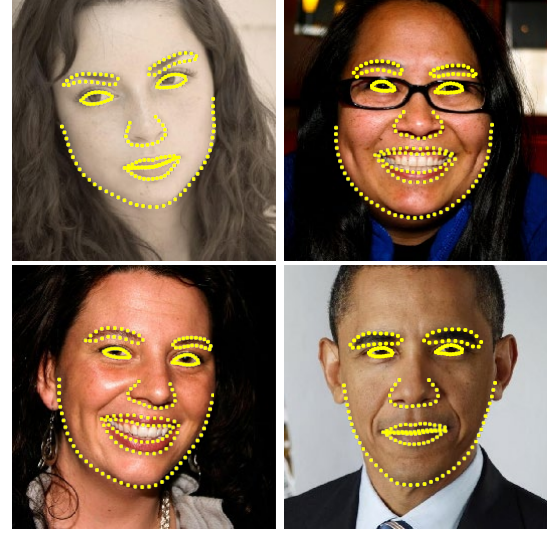
### 1. Giriş

Bu makalede, yüz hizalama işlemini milisaniyede gerçekleştiren ve standart veri setinde en son teknolojiye dayanan yöntemlerden daha üstün ya da bu yöntemlerle karşılaştırılabilecek derecede bir kesinlik elde eden yeni bir algoritma sunulmaktadır. Önceki yöntemlerle göre daha hızlı oluşu, önceki yüz hizalama algoritmalarının temel bileşenlerini tanımlamanın ve daha sonra bunların aerodinamik bir formül içinde, gradyan destekli yöntemle öğrenilen yüksek kapasiteli regresyon fonksiyonlarının kaskadına dahil edilmesinin bir sonucudur.

Biz de diğerlerinin yaptığı gibi [8, 2], yüz hizalamanın regresyon fonksiyonları kaskadıyla çözülebileceğini gösteriyoruz. Bizim incelediğimiz durumda, kaskadadaki her bir regresyon fonksiyonu, başlangıç tahmininden ve bu tahmine bağlı olarak indekslenen seyrek bir dizi pikselin yoğunluğu kullanılarak şekil etkin bir biçimde tahmin edilmektedir. Bizim çalışmamız, son on yılda yapılan ve yüz hizalamada önemli ilerleme sağlayan pek çok araştırmaya dayanmaktadır [9, 4, 13, 7, 15, 1, 16, 18, 3, 6, 19]. Özellikle, öğrendiğimiz regresyon fonksiyonlarına, atıf yapılan başarılı algoritmaların birçoğunda bulunan iki temel unsuru eklenmiştir ve halihazırda bu unsurları detaylandırılmıştır.

Şekil 1. HELEN veri setinde seçilen sonuçlar. Tek bir görüntüden yüzdeki 194 referans noktasını bir milisaniyede algılamak için, bir rasgele regresyon ağaçları dizisi kullanılmaktadır.

İlki, şeklin geçerli tahmini ile bağlantılı olarak, piksel yoğunluğunun indekslenmesi etrafında dönmektedir. Bir yüz görüntüsünün vektör sunumunda elde edilen detaylar hem şekil deformasyonu hem de aydınlatma koşullarındaki değişiklikler gibi sorun çıkaran faktörlere bağlı olarak büyük ölçüde değişebilir. Bu, bu detayların kullanıla



arak şeklin doğru tahmin edilmesini güçleştirmektedir. Bu durumda, hem şekli doğru bir biçimde tahmini için güvenilir detaylara ihtiyaç var; hem de, diğer taraftan, güvenilir detayları elde etmek için şeklin doğru tahminine ihtiyaç vardır. Diğer çalışmalar gibi [4, 9, 5, 8] bu çalışma da, bu sorunun üstesinden gelmek için yinelemeli bir yaklaşım (kaskad) kullanılmaktadır. Görüntünün global koordinat sisteminde çıkarılan detaylarına bağlı şekil parametrelerini geriletmek yerine, görüntü, şeklin mevcut tahmini esas alınarak normalleştirilmiş bir koordinat sistemine dönüştürülür ve daha sonra, şekil parametreleri için bir güncelleme vektörü tahmin etmek amacıyla detaylar çıkarılır. Bu işlem yakınsamaya kadar genellikle birkaç kez yinelenir.

İkincisinde ise, çıkarım/tahmin sorununun zorluğuyla nasıl mücadele edileceği değerlendirilir. Test esnasında, bir hizalama algoritması, şekli, yani görüntü verisine ve bizim model şeklimize en iyi uyan yüksek boyutlu bir vektörü tahmin etmek zorundadır. Sorun, birçok yerel tepe noktası olan dışbükey olmayan bir yapıdadır. Başarılı algoritmalar [4, 9], örneğin eğitim şekillerinin esas bileşenlerinin bulunması yoluyla, tahmini şeklin keşfedilebilecek doğrusal bir alt uzayda olduğu varsayımıyla, bu sorunun üstesinden gelmektedir. Bu varsayım, çıkarım sırasında göz önüne alınması gereken olası şekillerin sayısını büyük ölçüde azaltır ve yerel tepe noktasından kaçınmaya yardımcı olabilir. Son çalışma [8, 11, 2] belirli bir regresyon sınıfı eğitim şekilleriyle tanımlanan doğrusal bir alt uzayda yer alan tahminleri kesinlikle ürettiği ve ilave kısıtlamalara gerek olmadığı gerçeğini kullanır. Esasen, bizim regresyon fonksiyonlarımız bu iki öğeye sahiptir.

Bizim etkin regresyon fonksiyonu öğrenimi bu iki faktörle bir aradadır. Uygun bir kayıp fonksiyonu optimize edilmektedir ve detay (feature) seçimi işlemi veri odaklı bir şekilde yapılmaktadır. Özellikle, her bir regresörü, test zamanında minimize etmek istediğimiz aynı kayıp fonksiyonu olan karesel hata kaybı fonksiyonuyla gradyan destekli [10] yöntemle öğreniyoruz. Regresörün girişi olarak kullanılan seyrek piksel seti, gradyan destekli algoritma ile giriş pikselleri çiftleri arasındaki mesafenin önceki olasılıklarının bir kombinasyonu yoluyla seçilir. Önceki dağılım, destekli algoritmanın çok sayıda ilgili detayı keşfetmesini sağlar. Sonuç, ortalama yüz pozunu ile başlatıldığında, yüzdeki belirgin noktaların yerini belirleyebilen bir regresyon kaskadıdır.

Bu çalışmanın en büyük katkıları;

1. Regresyon ağaçları dizisine dayanan ve test esnasında minimize etmek istediğimiz eğitim süresince aynı kayıp fonksiyonunu minimize ederken, şekil değişikliği olmadan detay seçimini uygulayan yeni bir hizalama yöntemi.
2. Yöntemde, eksik veya belirgin olmayan etiketleri ele alan, doğal bir uzantı sunulmaktadır.
3. Yöntemimizin yüksek kaliteli tahminler ürettiğini ve kendisinden önceki en iyi yöntemden çok daha verimli olduğunu teyit eden nicel ve nitel sonuçlar sunulmuştur (Şekil 1).
4. Eğitim verilerinin miktarı, kısmen etiketlenmiş verilerin kullanımı ve sentezlenmiş verilerin tahminlerin kalitesi üzerindeki etkileri analiz edilmektedir.

## 2. Yöntem

Makalede, yüzdeki referans noktaları konumunu sayısal olarak verimli bir şekilde tam olarak tahmin etmek için bir algoritma sunulmaktadır. Önceki çalışmalar [8, 2] gibi, bizim çalışmamızda da bir regresörler kaskadından faydalanılmıştır. Bu bölümün geri kalanında, kaskadın münferit bileşenleri biçiminin ayrıntılarını ve eğitim uygulamamız açıklanmaktadır.

### 2.1. Regresörler Kaskadı

İlk olarak, bazı gösterimleri sunalım:  $x_i \in \mathbb{R}^2$  'dir ve bir  $I$  görüntüsündeki  $i$ 'inci yüz referans noktasının  $x$ ;  $y$  koordinatları olsun. O halde,  $S = (x^{T_1}, x^{T_2}, \dots, x^{T_p})^T \in \mathbb{R}^{2p}$  vektörü,  $I$  görüntüsündeki tüm  $p$  yüz referans noktalarının koordinatlarını belirtir. Bu çalışmada,  $S$  vektöründen sıklıkla, "şekil" olarak bahsedilecektir.  $S$ 'nin güncel tahminini  $\hat{S}^{(t)}$  ile gösterilir. Kaskaddaki her bir regresör  $r_t(\cdot, \cdot)$ , görüntü ve  $\hat{S}^{(t)}$  'yi kullanarak bir güncelleme vektörü tahmin eder ve bu tahmini geliştirmek için  $\hat{S}^{(t)}$  şekil tahminine eklenir.

$$\hat{S}^{(t+1)} = \hat{S}^{(t)} + r_t(I, \hat{S}^{(t)}) \quad (1)$$

Kaskadın kritik noktası şudur: regresör<sub>t</sub> tahminlerini piksel yoğunluk değerleri gibi detaylara bağlı olarak yapar,  $I$  kullanılarak hesaplanır ve  $\hat{S}^{(t)}$  güncel şekil tahminine göre indekslenir. Bu, sürece geometrik bir değişmezlik kazandırır ve kaskad ilerledikçe, yüzdeki hassas semantik bölgenin indekslendiğinden daha emin olunur. Daha sonra, bu indekslemenin nasıl yapıldığı açıklanacaktır.

Dizi tarafından genişletilen çıktı aralığının, ilk tahmin  $\hat{S}^{(0)}$  bu alana ait ise, eğitim verilerinin doğrusal bir alt uzayında yer alması sağlanır. Bu nedenle, yöntemimizi büyük ölçüde basitleştiren öngörüler üzerinde ilave kısıtlamalar uygulamak zorunda kalınmaz. İlk şekil, basitçe jenerik bir yüz algılayıcının çerçeve çıktısına göre ölçeklendirilen ve merkezlendirilen eğitim verisinin ortalama şekli olarak seçilebilir.

Her bir  $r_t$  'yi eğitmek için, [10]'da anlatıldığı gibi, gradyan ağacı destekli algoritma ve toplam karesel hata kaybını kullanıyoruz. Şimdi bu süreçle ilgili ayrıntıları paylaşalım.

### 2.2. Kaskaddaki Herbir Regresörün Öğrenilmesi

Her  $I_i$  'nin bir yüz görüntüsü ve  $S_i$  'nin bir şekil vektörü olduğu bir eğitim verisine sahip olduğumuzu varsayalım ( $I_i, S_i$ ). . . , ( $I_n, S_n$ ). Kaskaddaki ilk regresyon fonksiyonu olan  $r_0$ 'ı öğrenmek için eğitim verisinden bir yüz görüntüsü, bir şekil tahmini için başlangıç değeri ve hedef güncelleme adımı olmak üzere üçlü veri seti  $(I_{\pi_i}, \hat{S}_i^{(0)}, \Delta S_i^{(0)})$  oluşturalım, Burada

$$\begin{aligned} \pi_i &\in \{1, \dots, n\} \\ \hat{S}_i^{(0)} &\in \{S_1, \dots, S_n\} \setminus S_{\pi_i} \quad \text{ve,} \\ \Delta S_i^{(0)} &= S_{\pi_i} - \hat{S}_i^{(0)} \end{aligned}$$

$i = 1, \dots, N$  'dir.

Bu üçlülerin toplam sayısını  $N = nR$  olarak belirledik, burada  $R$ , her bir  $I_i$  görüntüsü için kullanılan başlatma sayısıdır. Her bir görüntünün başlangıçtaki tahmini şekli,  $\{S_1, \dots, S_n\}$  kullanılarak yer değiştirmeden uniform olarak örneklendirilir.

Bu veriden karesel hata kayıplarının toplamı ile gradyan destekli regresyon ağaçlarını kullanarak  $r_0$  regresyon fonksiyonunu (bkz. Algoritma 1) öğrenmekteyiz. Daha sonra  $(I_{\pi_i}, \hat{S}_i^{(1)}, \Delta S_i^{(1)})$  eğitim verisini elde etmek maksadıyla, kaskaddaki bir sonraki regresör olan  $r_1$ , ( $t=0$ ) yapılarak, üçlü eğitim seti güncellenmektedir.

$$\hat{S}_i^{(t+1)} = \hat{S}_i^{(t)} + r_t(I_{\pi_i}, \hat{S}_i^{(t)}) \quad (5)$$

$$\Delta S_i^{(t+1)} = S_{\pi_i} - \hat{S}_i^{(t+1)} \quad (6)$$

Bu işlem birleştirildiğinde yeterli düzeyde doğruluk seviyesi sağlayan, T tane regresörü olan  $(r_0, r_1, \dots, r_{T-1})$  bir kaskad öğrenilinceye kadar yinelenir.

Daha önce belirtildiği gibi, her bir  $r_t$  regresörü, gradyan destekli regresyon ağaçların algoritması kullanılarak öğrenilir. Unutulmamalıdır ki, karesel hata kaybı kullanılmakta ve en içteki döngüde hesaplanan artık değerler, her bir eğitim örneğinde değerlendirilen bu kayıp fonksiyonunun gradyanına karşılık gelmektedir. Algoritma ifadesinde yer alan öğrenme hızı parametresi  $(0 < \nu < 1)$  aynı zamanda daralma (shrinkage) faktörü olarak da bilinir.  $\nu < 1$  olarak ayarlandığında, aşırı uyumla mücadeleye yardımcı olur ve genellikle  $\nu = 1$  olduğunda öğrenilenlerden daha iyi genelleştiren regresyonlara yol açar [10].

### Algoritma 1 Kaskaddaki $r_t$ 'nin öğrenilmesi

Eğitim verisi  $\{(I_{\pi_i}, \hat{S}_i^{(t)}, \Delta S_i^{(t)})\}_{i=1}^N$  olarak ve öğrenme hızı (daralma faktörü) da  $(0 < \nu < 1)$  olarak alınır.

1. Başlat

$$f_0(I, \hat{S}^{(t)}) = \arg \min_{\gamma \in \mathbb{R}^{2p}} \sum_{i=1}^N \|\Delta S_i^{(t)} - \gamma\|^2$$

2.  $k=1, \dots, K$  için

- (a)  $i=1, \dots, N$  için

$$r_{ik} = \Delta S_i^{(t)} - f_{k-1}(I_{\pi_i}, \hat{S}_i^{(t)})$$

- (b)  $g_k(I, \hat{S}^{(t)})$  zayıf regresyon fonksiyonunu vererek hedef  $r_{ik}$ 'lara regresyon ağacının uydurulması

- (c) Güncelleme

$$f_k(I, \hat{S}^{(t)}) = f_{k-1}(I, \hat{S}^{(t)}) + \nu g_k(I, \hat{S}^{(t)})$$

3. Çıktı  $r_t(I, \hat{S}^{(t)}) = f_K(I, \hat{S}^{(t)})$

### 2.3. Ağaca bağlı regresör

Her bir  $r_t$  regresyon fonksiyonunun özü, ağaca bağlı regresörlerin, gradyan destekli algoritma boyunca, artık değer hedeflerine uyumudur. Şimdi, her bir regresyon ağacının çalışılması için en önemli uygulama ayrıntılarını gözden geçirecektir.

#### 2.3.1 Şekli değişmeyen bölünmüş testler

Regresyon ağacındaki her bölünme düğümünde (split node), iki piksel yoğunluğu arasındaki farkın eşik değerine bağlı bir karar verilmektedir. Testte kullanılan pikseller ortalama şeklin koordinat sisteminde tanımlanan  $u$  ve  $v$  konumlarında bulunmaktadır. Rasgele şekli olan bir yüz görüntüsü için, şekle bağlı aynı  $u$  ve  $v$  koordinatlarına sahip noktaların, ortalama şekle göre indekslenmesi istenmektedir. Bunu başarmak için, detayların elde edilmesinden önce görüntü, mevcut tahmini şekle dayalı olarak ortalama şekle göre deforme (warp) edilebilir. Görüntünün sadece çok seyrek bir temsilini kullandığımızdan, tüm görüntünün aksine noktaların yerinin deforme edilmesi çok daha etkilidir. Dahası, kaba bir deformasyon yaklaşımı yalnızca [2]'de önerilen yerel dönüşümlere ilaveten yapılacak genel benzerlik dönüşümüyle gerçekleştirilebilir.

Net ayrıntılar aşağıdaki gibidir.  $k_u$ , ortalama şekilde  $u$ 'ya en yakın olacak şekilde yüz üzerindeki bir noktanın indeksi olsun ve bu noktanın  $u$ 'dan olan ötelemesi (offset) şu şekilde tanımlansın,

$$\delta x_u = u - \bar{x}_{k_u}$$

Daha sonra  $I_i$  görüntüsünde tanımlanan  $S_i$  şekli için, ortalama şekildeki  $u$ 'ya nitelik olarak benzer olan  $I_i$ 'deki konum, şu şekilde verilmektedir:

$$u' = x_{i, k_u} + \frac{1}{s_i} R_i^T \delta x_u \quad (7)$$

Burada,  $s_i$  ve  $R_i$ ,  $S_i$ 'yi ortalama şekil olan  $\bar{S}$ 'ye dönüştüren benzerlik dönüşümünün ölçek ve dönüklük matrisidir. Ölçek ve dönüklüğün ortalama şekilde  $\bar{X}_j$  yüzdeki referans noktalar (landmark) ile deforme edilen şekildeki noktalar arasındaki farkın kareleri toplamını minimize ettiği tespit edilmiştir.  $v'$  de benzer şekilde tanımlanır.

$$\sum_{j=1}^p \|x_j - (s_i R_i x_{i,j} + t_i)\|^2 \quad (8)$$

Normal olarak her bir bölüm noktasında üç parametrelili bir karar vardır,  $\theta = (\tau, u, v)$ . Bu aşağıdaki şekilde her bir eğitim ve test örneğine uygulanmaktadır:

$$h(I_{\pi_i}, \hat{S}_i^{(t)}, \theta) = \begin{cases} 1 & I_{\pi_i}(u') - I_{\pi_i}(v') > \tau \\ 0 & \text{diğer durumda} \end{cases} \quad (9)$$

burada  $u'$  ve  $v'$ , (7) numaralı eşitliğe göre  $\hat{S}_i^{(t)}$ 'yi  $\bar{S}$ 'ye en iyi şekilde deforme eden ölçek ve dönüklük matrisi kullanılarak belirlenir.

Uygulamada atanan değerler ve yerel dönüşümler çalışma aşamasında belirlenir. Test zamanında bu işlemin bilgisayarı en çok yoran bölümü olan bu benzerlik dönüşümünün hesaplanması, yalnızca kaskadın her bir seviyesinde yapılır.

### 2.3.2 Dügüm bölmelerini seçme

Her bir regresyon ağacı için, her bir yaprak düğüme sabit vektörü uyduran parçalı sabit fonksiyonla temel fonksiyona yaklaşıyoruz. Regresyon ağacını eğitmek için, her bir düğümden rasgele bir dizi aday bölünme, yani  $\theta^*$ 'lar üretilir. Daha sonra hırsla, bu adaylardan karesel hata toplamını minimize edecek  $\theta^*$  değerleri seçilir. Eğer  $Q$ , bir düğümden eğitim örneklerinin indislerinin kümesiyse bu aşağıdaki eşitliğin minimize edilmesine karşılık gelir:

$$E(Q, \theta) = \sum_{s \in \{l, r\}} \sum_{i \in Q_{\theta, s}} \|r_i - \mu_{\theta, s}\|^2 \quad (10)$$

Burada,  $Q_{\theta, l}$ ,  $\theta$  tarafından tetiklenen karar nedeniyle, sol düğüme gönderilen örneklerin indisleridir,  $r_i$ , gradyan destekli algoritmada i görüntüsü için hesaplanan tüm artık değerlerin (residual) vektörüdür ve

$$\mu_{\theta, s} = \frac{1}{|Q_{\theta, s}|} \sum_{i \in Q_{\theta, s}} r_i, \quad \text{for } s \in \{l, r\} \quad (11)$$

Optimal bölünme çok etkin görünebilir, çünkü eşitlik (10) yeniden düzenlenirse ve  $\theta$ 'ya bağlı olmayan faktörler ihmal edilirse aşağıdaki eşitlik elde edilir:

$$\arg \min_{\theta} E(Q, \theta) = \arg \max_{\theta} \sum_{s \in \{l, r\}} |Q_{\theta, s}| \mu_{\theta, s}^T \mu_{\theta, s}$$

Burada, farklı  $\theta$ 'ları değerlendirirken yalnızca  $\mu_{\theta, l}$ 'yi hesaplamamız gerekir. Çünkü  $\mu_{\theta, r}$ , ana düğüm olan  $\mu$ 'deki ve  $\mu_0$ 'deki hedeflerin ortalaması kullanılarak aşağıdaki şekilde hesaplanabilir:

$$\mu_{\theta, r} = \frac{|Q| \mu - |Q_{\theta, l}| \mu_{\theta, l}}{|Q_{\theta, r}|}$$

### 2.3.3 Detay seçimi

Her düğümden karar, bir çift pikseldeki yoğunluk değerleri farkının eşiklenmesine dayanmaktadır. Bu oldukça basit bir testtir, ancak küresel aydınlatmada meydana gelen değişikliklere karşı hassas olmaması nedeniyle, tek yoğunluklu eşiklemeden çok daha güçlüdür. Fakat, piksel farklarını kullanmanın dezavantajı, potansiyel bölünme (detay) adaylarının sayısının ortalama görüntüdeki piksel sayısının ikinci dereceden olmasıdır. Bu da, çok fazla sayıda arama yapmadan iyi  $\theta$ 'ların bulunmasını zorlaştırmaktadır. Bununla birlikte, bu sınırlayıcı faktör, görüntü verisinin yapısı hesaba katılarak bir dereceye kadar hafifletilebilir. Daha yakın piksel çiftlerinin seçilmesinin teşvik edilmesi için, bölme işleminde kullanılan pikseller arasındaki mesafenin üstel önceliği aşağıdaki şekilde verilmektedir.

$$P(\mathbf{u}, \mathbf{v}) \propto e^{-\lambda \|\mathbf{u} - \mathbf{v}\|} \quad (12)$$

Bir dizi yüz veri setinde, bu basit önceliklendirmeyi kullanmanın, tahmin hatasını azalttığı belirlenmiştir. Şekil 4'de, bu önceliğin kullanılması ve kullanılmaması durumlarında, detay havuzunun 20'ye sabitlenmesi ile seçilen detaylar karşılaştırılmaktadır.

### 2.4. Eksik etiketlerin ele alınması

Eşitlik (10)'un amacı, eğitim görüntülerinin bazılarında referans noktaların bazılarının etiketlenmediği (veya her bir nokta için bir belirsizlik ölçüsü vardır) durumunu ele almak üzere kolaylıkla genişletilebilir. Her bir  $i$  eğitim görüntüsü ve her bir  $j$  referans noktası için,  $w_{i,j} \in [0, 1]$  değişkenleri tanımlanmaktadır.  $W_{i,j}$  değerine 0 atanması,  $j$  referans noktasının  $i$ 'inci görüntüde etiketlenmediğini gösterirken bu değer 1 olması görüntüde etiketlendiğini gösterir. Daha sonra eşitlik (10) aşağıdaki şekilde güncellenir,

$$E(Q, \theta) = \sum_{s \in \{l, r\}} \sum_{i \in Q_{\theta, s}} (\mathbf{r}_i - \mu_{\theta, s})^T W_i (\mathbf{r}_i - \mu_{\theta, s})$$

Burada  $W_i$  diyagonal elemanları ( $w_{i1}, w_{i1}, w_{i2}, w_{i2}, \dots, w_{ip}, w_{ip}$ )<sup>T</sup> vektörü olan diyagonal bir matristir.

$$\mu_{\theta, s} = \left( \sum_{i \in Q_{\theta, s}} W_i \right)^{-1} \sum_{i \in Q_{\theta, s}} W_i \mathbf{r}_i, \quad \text{for } s \in \{l, r\} \quad (13)$$

Gradyan destekli algoritma da bu ağırlık faktörlerini dikkate alacak şekilde değiştirilmelidir. Dizi modeli hedeflerin ağırlıklı ortalaması ile başlatılarak ve regresyon ağaçları algoritma l'deki ağırlıklı artık değerlere aşağıdaki gibi uydurularak bu gerçekleştirilebilir.

$$\mathbf{r}_{ik} = W_i (\Delta S_i^{(k)} - f_{k-1}(I_{\pi_i}, \hat{S}_i^{(k)})) \quad (14)$$

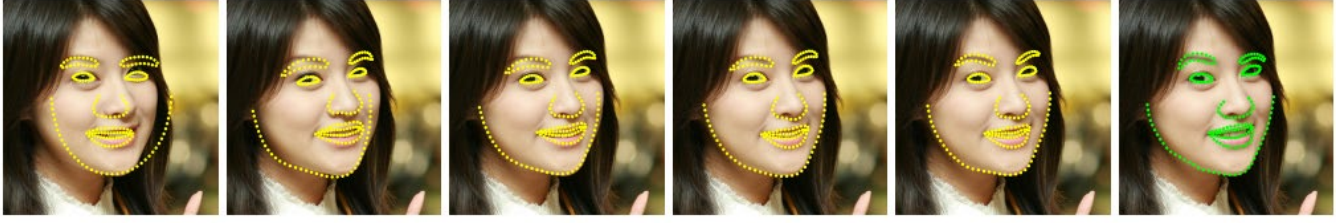
### 3. Deneyler

**Ana hata:** Performansın doğru bir şekilde ölçmek için, önerdiğimiz yöntem olan bir regresyon ağacı dizisinde (ERT), iki temel hat daha oluşturulmuştur. Birincisi, rasgele detay seçimli (EF) rasgele ferni esas almaktadır ve diğeri ise bunun daha gelişmiş bir versiyonu olan ve [2]'yi yeniden uygulayarak gerçekleştirilen korelasyon tabanlı detay seçimidir (EF+CB). Tüm parametreler her üç yaklaşım için sabittir.

EF, dizi içinde rasgele fernlerin zayıf regresörler olarak ileri yönlü uygulanmasını kullanır ve bu en hızlı denenen yoldur. Fernleri düzenli hale getirmek (regularize) için [2]'de önerilen aynı daralma yöntemi kullanılmaktadır.

EF + CB,  $r_i$  hedef çıktılarını rastgele bir  $w$  yönü üzerinde tahmin eden korelasyon temelli bir detay seçim yöntemidir ve  $(u, v)$  s.t. detay çiftlerini seçer. Tahmini hedefleri  $\mathbf{w}^T \mathbf{r}_i$  olan eğitim verisi üzerinde en yüksek örneklem korelasyonuna  $I_i(u_0) - I_i(v^*)$  sahiptir.

**Parametreler:** Belirtilmediği sürece, tüm deneyler aşağıdaki sabit parametrelerle gerçekleştirilmiştir. Kaskaddaki güçlü regresörlerinin ( $r_i$ ) sayısı  $T = 10$ , ve her bir  $r_i$ ,  $K = 500$  tane zayıf regresör ( $g_k$ ) içermektedir.  $g_k$ 'yi temsil etmek için kullanılan ağaçların (veya fernlerin) derinliği  $F = 5$  olarak ayarlanır. Kaskadın her seviyesinde, görüntüden  $P=400$  piksel konumu örneklenir. Zayıf regresörleri eğitmek için, bu  $P$  piksel koordinat çiftinden önceliğe göre rasgele bir örnek alınmaktadır ve zayıf regresörleri eğitmek için, bu  $P$  piksel koordinatlarından bir çifti rastgele örneklemeye tabi tutulur ve eşitlik (9)'da açıklandığı gibi potansiyel bölünme oluşturmak için rastgele bir eşik değeri seçilir. Daha sonra en iyi bölünme, bu işlem  $S = 20$  kez yinelenerek ve hedefimizi optimize eden bölünme seçilerek bulunur. Modelimizin öğrenilmesi için eğitim verilerinin oluşturulmasında her eğitim örneği için  $R = 20$  farklı ilk değer kullanılmaktadır.



(a) T = 0 (b) T = 1 (c) T = 2 (d) T = 3 (e) T = 10 (f) Kesin referans

Şekil 2. Temel Viola & Jones [17] yüz tespit çıktısını merkezleyen ortalama şekle göre başlatılan kaskadın farklı seviyelerindeki referans nokta tahminleri. Kaskadın ilk seviyesinden sonra hata zaten çok azalmış durumdadır.

**Performans:** Algoritmanın tek bir görüntü üzerindeki çalışma zamanı karmaşıklığı sabittir  $O$  (TKF). Eğitim süresinin karmaşıklığı, doğrusal olarak eğitim verisinin  $O$ (NDTKFS) sayısına bağlıdır, burada  $N$ , eğitim verisinin sayısı ve  $D$ , hedeflerin boyutudur. Tek bir CPU ile yapılan uygulamada, algoritmamızla, HELEN [12] veri setinde yapılan eğitim yaklaşık bir saat almaktadır ve çalışma anında her bir resim başına sadece yaklaşık bir milisaniye sürmektedir.

**Veritabanı:** Bildirilen deneysel sonuçların çoğu, en zorlu açık kaynaklı veri seti olarak gördüğümüz HELEN [12] yüz veri tabanı için yapılmıştır. Toplam 2330 görüntüden oluşmakta ve bunların her biri 194 referans nokta belirtmektedir. Yazarların önerdiği şekilde, 2000 görüntü eğitim verisi olarak ve diğerleri ise test için kullanılmıştır.

Ayrıca, 1432 görüntüden oluşan popüler LFPW [1] veritabanına ilişkin nihai sonuçlar da bildirilmektedir. Ne yazık ki, 778 eğitim görüntüsü ve 216 geçerli test görüntüsü indirebilmiştir. Bu da sonuçlarımızın bu veri setine ilişkin daha önceden bildirilen değerlerle doğrudan karşılaştırılamamasına neden olmuştur.

**Karşılaştırma:** Tablo 1’de, sonuçlarımızın önceki algoritmalarla özet bir karşılaştırması vardır. Ana hatlarımıza ek olarak, sonuçlar Aktif Şekil Modelinin iki çeşidi olan STASM [14] ve CompASM [12] ile karşılaştırılmıştır.

	[14]	[12]	EF	EF+CB	EF+CB (5)	EF+CB (10)	ERT
Hata	,111	,091	,062	,059	,059	,055	<b>,049</b>

Tablo 1. Farklı algoritmaların HELEN veri kümesindeki sonuçlarının özeti. Hata, her bir konum işaretinin kesin referans konumuna olan normalize edilmiş ortalama uzaklığıdır. Mesafeler, interküler mesafeye bölünerek normalize edilir. Köşeli parantez içindeki sayı, regresyon algoritmasının rastgele başlangıç noktasıyla kaç kez yinlendiğini gösterir. Hiçbir rakam yoksa, yöntem ortalama şekil ile başlatılmıştır. Birden fazla tahmin söz konusu olduğunda, tahminlerin ortanca değeri referans noktasının nihai tahmini olarak seçilmiştir.

Bu makalede açıklanan regresyon ağaçları dizisi, fern dizisi üzerinde sonuçları önemli ölçüde iyileştirmektedir. Şekil 3, kaskadın farklı seviyelerindeki ortalama hatayı göstermektedir; bu, ERT'nin hatayı diğer ana hatlara göre çok daha hızlı bir şekilde azaltabileceğini göstermektedir. EF+CB'nin birden çok kere çalıştırılmasıyla elde edilen sonuçları ve nihai tahminlerin ortanca değeri de sunulduğuna dikkat edilmelidir. Sonuçlar, metodumuzu kullanarak, daha az miktarda hesaplama ile, EF+CB'ye yakın hata oranının elde edilebileceğini göstermektedir.

Ayrıca yaygın olarak kullanılan LFPW [1] veri setine ilişkin elde edilen sonuçlar da sunulmaktadır (Tablo 2). EF+CB ana hattı ile [2]'de bildirilen sayılar çoğaltılmamıştır. (Bunun sebebi, veri setinin tamamının elde edilememesi olabilir.) Bununla birlikte, yöntemimiz, bu veri setine ilişkin daha önce bildirilen sonuçların çoğunu aşmakta ve diğer yöntemler için gereken hesaplama süresinin yalnızca bir kısmı kadar sürmektedir.

	[1]	[2]	EF	EF+CB	EF+CB (5)	EF+CB (10)	ERT
Hata	,040	<b>,034</b>	,051	,046	,043	,041	,038

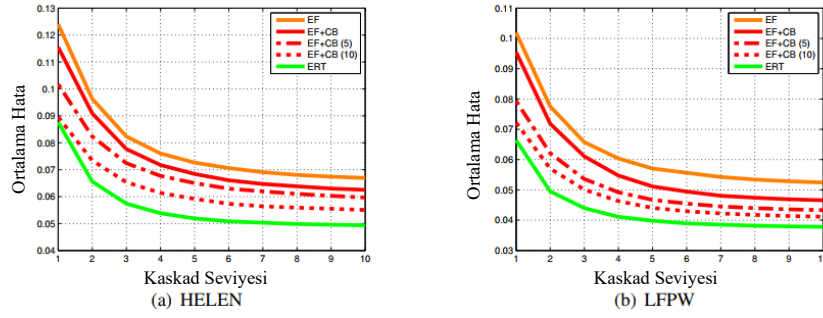
Tablo 2. LFPW veri setine uygulanan farklı yöntemlerin karşılaştırması. Rakamların açıklaması için lütfen Tablo 1'in başlığına bakınız.

**Detay Seçimi:** Tablo 4, bölünmede kullanılan pikseller arasındaki mesafede, uniform öncelik yerine eşitlik (12)'nin öncelik olarak kullanılmasının nihai sonuçlar üzerindeki etkisini göstermektedir.  $\lambda$  parametresi, detaylarımızdaki iki piksel arasındaki etkin maksimum mesafeyi belirler ve deneylerimizde 0.1 olarak ayarlanmıştır. Kaskaddaki her bir güçlü regresörün ( $r$ ) öğrenilmesinde bu parametrenin çapraz doğrulama ile seçilmesi daha belirgin bir iyileştirme potansiyeli sağlar. Şekil 4'te farklı öğeler kullanıldığında seçilen detay çiftleri gösterilmektedir.

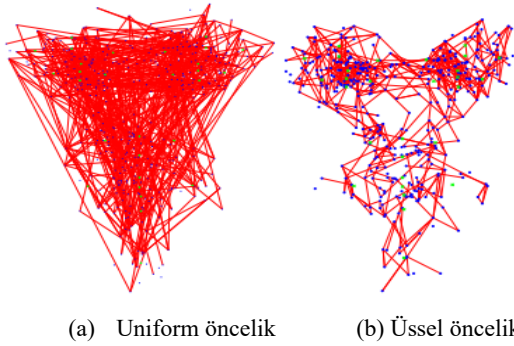
	Uniform	Üssel
Hata	,053	<b>,049</b>

Tablo 3. Nihai ortalama hata üzerinden detay seçimi için farklı öncelik kullanılmasının etkisi. Bir detayı belirleyen iki piksel arasındaki Öklid uzunluğuna üssel bir öncelik uygulanmaktadır, bkz. Eşitlik (12).

**Düzenleme (Regularization):** Gradyan destekli algoritma kullanılırken aşırı uyumdan kaçınmak için dikkatli olunması gerekir. Test hatalarını düşürmek için, çeşitli düzenlemelerin yapılması gerekmektedir. En basit yaklaşım daraltmadır (sshrinkage).



Şekil 3. HELEN (a) ve LFPW (b) veri kümelerinde uygulanan farklı yöntemlerin karşılaştırması. EF, rasgele fernlerin dizisidir ve EF+CB, ortalama şekil ile başlatılmış korelasyon temelli detay seçimine ilişkin fernlerin dizisidir. [2]'de önerildiği gibi, çeşitli başlangıç değerlerine (5 ve 10) ilişkin sonuçların ortanca değerinin alınmasıyla bulunan sonuçlar da sunulmaktadır. Sonuçlar, başlangıç değeri olarak sadece ortalama şekli kullanan önerilen regresyon ağacının (ERT), sürekli olarak fern ana hatları dizisinin daha iyi performans sergilediğini göstermektedir ve daha az hesaplama ile aynı hata oranına ulaşabilmektedir.



Şekil 4. Farklı öncelikler kullanıldığında farklı detaylar seçilmektedir. Üssel öncelik, seçimi birbirine yakın olan piksel çiftlerine doğru yönlendirir.

Bu, gradyan yükseltme algoritmasında öğrenme hızının ( $\nu$  değerinin) 1'in altına düşürülmesini gerektirir (burada,  $\nu = 0,1$  olarak alınmaktadır). Düzenleme, çoklu regresyon ağaçlarının tahminlerinin ortalaması alınarak da gerçekleştirilebilir. Bu şekilde,  $g_k$ , bir ağaç yerine rastgele bir ormana karşılık gelir ve  $\nu = 1$  olarak ayarlanır. Bu nedenle, bir regresyon ağacının artık değerlere uyması yerine, gradyan destekli algoritmanın her yinelemesinde, birden çok ağaca (deneylerimizde 10) uydurulmaktadır ve sonuçların ortalaması alınmaktadır. (Toplam ağaç sayısı her durumda sabittir.)

Yanlılık ve varyans dengesi açısından, gradyan destekli algoritma her zaman yanlılığı azaltır, ancak varyansı da artırır. Fakat daralma veya etkin ortalamayla yapılan düzenleme, örtüşen birden çok modelin öğrenilmesiyle varyansı azaltır.

	Düzenlenmemiş	Daralma	Ortalama
Hata	,103	,049	,049

Tablo 4. Farklı düzenleme yöntemleri uygulandığında HELEN veri setinde görülen sonuçların karşılaştırması. Dizideki toplam ağaç sayısı aynı olduğunda, daralma veya ortalamanın kullanımında benzer sonuçlar elde edilmiştir.

Daha standart daralma yöntemiyle kıyaslandığında, ortalama alınarak yapılan düzenleme kullanılarak da benzer sonuçlar elde

edilmiştir. Bununla birlikte, ortalamaya göre düzenleme, daha ölçeklenebilir olma avantajına sahiptir, bu eğitim süresi boyunca paralel işlemi mümkün kıldığı için, özellikle büyük ölçekli problemlerin çözümünde önemlidir.

**Kaskad:** Kaskadın her seviyesinde, ikinci düzey regresörler sadece şekil indeksli detayların sabit ve aralıklı bir alt kümesini gözlemleyebilir. Mevcut tahminlere dayanan detayların indekslenmesi, görüntüyü küçük bir maliyetle deforme etmenin (warp) basit bir yoludur. Tablo 5'de, kaskadın kullanılıp kullanılmaması durumuna göre nihai hata oranı gösterilmektedir. Bu yinelemeli mekanizmayı kullanarak, daha önce bildirilen sonuçlara paralel olarak anlamlı iyileşmeler saptanmıştır [8, 2] (Burada adil bir karşılaştırma yapılması için, gözlemlenen özelliklerin toplam sayısı  $10 \times 400$  noktayla sabitlenmiştir.)

Ağaç Sayısı	1x500	1x5000	10x500
Hata	,085	,074	,049

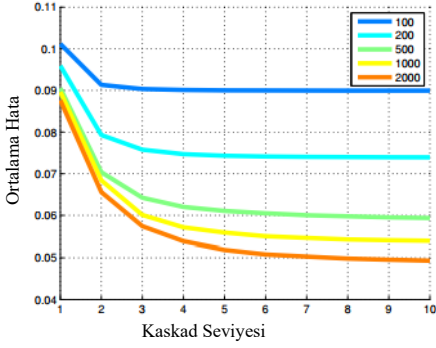
Tablo 5. Yukarıdaki sonuçlar, tek seviyeli bir dizi yerine regresyon kaskadının kullanılmasının önemini göstermektedir.

**Eğitim Verileri:** Eğitim görüntülerinin sayısına göre yöntemimizin performansını test etmek için, eğitim verilerinin farklı büyüklükteki alt gruplarından farklı modeller hazırlanmıştır. Tablo 6'da, nihai sonuçlar özetlenmektedir ve Şekil 5'te, kaskadın her seviyesindeki hatanın grafiği çizdirilmiştir. Çok sayıda eğitim örneği olduğunda, regresörlerin çok sayıda seviyesinin kullanımı, en kullanışlı yöntemdir.

Aynı deneyler, zenginleştirilmiş (augmented) örneklerin toplam sayısını sabit tutarak tekrarlanmıştır, ancak bir etiketli yüz örneğinden eğitim örneği oluşturmak için kullanılan başlangıç şekillerinin kombinasyonunu ve kaskadın öğrenilmesinde kullanılan etiketlenmiş görüntülerin sayısı çeşitlendirilmiştir (Tablo 7).

Örnek Sayısı	100	200	500	1000	2000
Hata	,090	,074	,059	,054	,049

Tablo 6. Eğitim örneklerinin sayısına göre nihai hata oranı. Kaskadlı regresörlerin öğrenilmesi için eğitim verisi oluştururken, yüzün başlangıç tahmini için, her bir etiketli yüz görüntüsü 20 farklı etiketli yüz kullanarak 20 eğitim örneği oluşturulmuştur.



Şekil 5. Kaskadın her seviyesindeki ortalama hata, kullanılan eğitim örneklerinin sayısına göre çizilmiştir. Eğitim örneklerinin sayısının çok olduğu durumlarda, çok sayıda regresör seviyesi kullanılması çok kullanışlıdır.

Örnek Sayısı	100	200	500	1000	2000
Başlangıç Şekil Sayısı	400	200	80	40	20
Hata	,062	,057	,054	,052	,049

Tablo 7. Burada eğitim örneklerinin etkin sayısı sabittir, ancak her bir etiketli yüz görüntüsü için kullanılan eğitim görüntülerinin sayısı, başlangıç şekillerinin sayısının farklı kombinasyonları kullanılmaktadır.

Eğitim verilerinin farklı başlangıç biçimleriyle artırılması, veri setini şekil bakımından genişletmektedir. Bulgularımız, bu tür bir zenginleşmenin, etiketli eğitim görüntüsü eksikliğini tam olarak telafi etmediğini göstermektedir. Buna rağmen, eğitim görüntülerinin sayısını artırarak elde edilen iyileşme oranı, ilk birkaç yüz görüntünün ardından hızla yavaşlamaktadır.

**Kısmi etiketlemeler:** Tablo 8, kısmen etiketlenmiş verilerin kullanılmasına ilişkin sonuçları göstermektedir. 200 eğitim örneği tamamen etiketlenmiştir ve geriye kalanlara sadece kısmen etiketlenmiştir.

Örnek Sayısı	200	200+1800 (25%)	200+1800 (50%)	1000	2000
Hata	,062	,057	,054	,052	,049

Tablo 8. Kısmen etiketli verilerin kullanıldığında elde edilen sonuçlar. 200 örnek her zaman tamamen etiketlidir. Parantez içindeki değerler, ölçü yapılan yer işaretlerinin yüzdesini göstermektedir.

Sonuçlar, kısmen etiketli verilerin kullanılmasıyla önemli iyileşme sağlanabileceğini göstermektedir. Ancak, görüntülenen iyileşme doymuş olmayabilir, çünkü şekil parametrelerinin boyutunun yer işaretlerinin boyutundan (194×2) çok daha düşük olduğu bilinmektedir. Bu nedenle, referans noktalarının koordinatları arasındaki korelasyon avantajından açık bir şekilde yararlanarak,

kısmi etiketlerle daha belirgin iyileştirme sağlanabilmesi potansiyeli bulunmaktadır. Bu makalede anlatılan gradyan destekli yöntemin referans noktaları arasındaki korelasyondan yararlanmadığına dikkat etmek gerekir. Bu konu gelecekteki bir çalışmada incelenebilir.

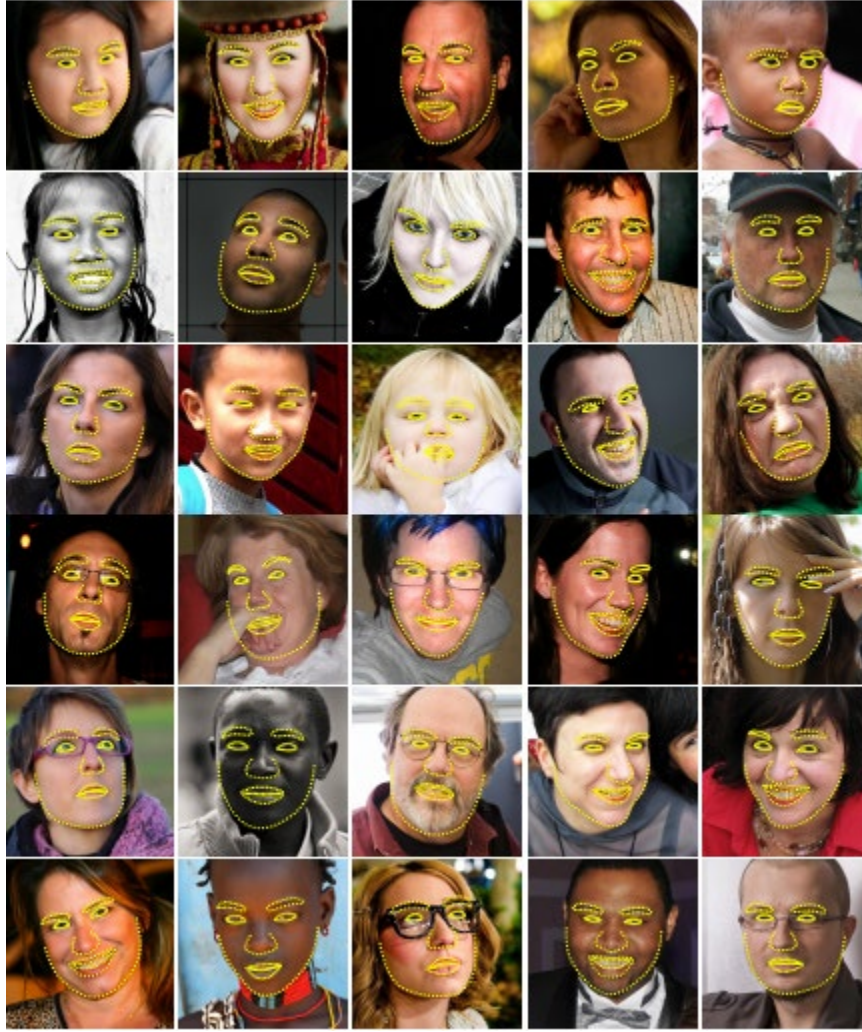
#### 4. Sonuç

Bir girdi resminden elde edilen yoğunluk değerlerinin seyrek bir alt kümesinden, yüze ilişkin referans noktalarının koordinatlarının kestirilmesi için nasıl bir regresyon ağacı dizisinin kullanılacağı açıklanmıştır. Sunulan çerçeve, önceki çalışmalara kıyasla hatayı daha kısa sürede azaltmaktadır ve kısmi veya belirsiz etiketleri de işleyebilmektedir. Algoritmamızın ana bileşenleri farklı hedef boyutlarını bağımsız değişken olarak görürken, bu çalışmanın doğal bir uzantısı olarak, daha etkin bir çalışma için şekil parametrelerinin korelasyonundan yararlanılması ve kısmi etiketlerin daha iyi kullanımı olabirdi.

**Teşekkür:** Bu çalışma VINST projesi kapsamında İsveç Stratejik Araştırma Vakfı tarafından finanse edilmiştir.

#### Kaynaklar:

- [1] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, ve N. Kumar. Örneklerde uzlaşma sağlayarak yüz parçalarının yerleştirilmesi. CVPR, sayfa 545–552, 2011. 1, 5
- [2] X. Cao, Y. Wei, F. Wen, ve J. Sun. Açık şekil regresyonuyla yüz hizalama. CVPR, sayfa 2887–2894, 2012. 1, 2, 3, 4, 5, 6
- [3] T. F. Cootes, M. Ionita, C. Lindner, ve P. Sauer. Rasgele Forest regresyon seçimi kullanılarak sağlam ve doğru şekil modeli uydurma. ECCV, 2012. 1
- [4] T. F. Cootes, C. J. Taylor, D. H. Cooper, ve J. Graham. Aktif şekil modelleri-eğitim ve uygulamaları. Computer Vision and Image Understanding, 61(1):38–59, 1995. 1, 2
- [5] D. Cristinacce and T. F. Cootes. Destekli regresyon aktif şekil modelleri. In BMVC, sayfa 79.1–79.10, 2007. 1
- [6] M. Dantone, J. Gall, G. Fanelli, ve L. V. Gool. Koşullu forest regresyonu kullanarak yüz detaylarının gerçek zamanlı tespiti. CVPR, 2012. 1
- [7] L. Ding ve A. M. Mart'inez. Yüzlerin ve yüz detaylarının ayrıntılı şekilde hassas tespiti. CVPR, 2008. 1
- [8] P. Dolla'r, P. Welinder, ve P. Perona. Kademeli poz regresyonu. In CVPR, sayfa 1078–1085, 2010. 1, 2, 6
- [9] G. J. Edwards, T. F. Cootes, ve C. J. Taylor. Aktif görünüş modellerindeki gelişmeler. ICCV, sayfa 137–142, 1999. 1, 2
- [10] T. Hastie, R. Tibshirani, ve J. H. Friedman. İstatistiksel öğrenmenin elemanları: veri madenciliği, çıkarım ve tahmin. New York: Springer-Verlag, 2001. 2, 3
- [11] V. Kazemi ve J. Sullivan. Parça tabanlı modellemeyle yüz hizalama. BMVC, sayfa 27.1–27.10, 2011. 2
- [12] V. Le, J. Brandt, Z. Lin, L. D. Bourdev, ve T. S. Huang. İnteraktif yüz detayı yerleştirilmesi. ECCV, sayfa 679– 692, 2012. 5
- [13] L. Liang, R. Xiao, F. Wen, ve J. Sun. Bileşene dayalı ayırt edici arama yoluyla yüz hizalama. ECCV, sayfa 72–85, 2008. 1



Şekil 6: HELEN veri tabanındaki nihai sonuçlar

- [14] S. Milborrow ve F. Nicolls. Genişletilmiş aktif şekil modeliyle yüz detaylarının tespiti. *ECCV*, sayfa 504–513, 2008. 5
- [15] J. Saragih, S. Lucey, ve J. Cohn. Düzenlenmiş referans noktası ortalama kaymalarıyla deforme edilebilir model uydurma. *Internation Journal of Computer Vision*, 91:200–215, 2010. 1
- [16] B. M. Smith ve L. Zhang. Parametrik olmayan şekil modelleriyle birleşik yüz hizalama. *ECCV*, sayfa 43–56, 2012. 1
- [17] P. A. Viola ve M. J. Jones. Gerçek zamanlı sağlam yüz tespiti. In *ICCV*, sayfa 747, 2001. 5
- [18] X. Zhao, X. Chai, ve S. Shan. Birleşik yüz hizalama: Düzenlenmiş yeniden uyumlarla, kötü hizalıların iyi olanlarla kurtarılması. *ECCV*, 2012. 1
- [19] X. Zhu ve D. Ramanan. Vahşi ortamda yüz tanıma, duruş tahmini ve referans noktası yerelleştirme. *CVPR*, sayfa 2879–2886, 2012. 1