

Projekat na kursu Mašinsko učenje

# Prepoznavanje poznatih ličnosti korišćenjem neuronskih mreža

Vuk Stefanović 66/2019  
Dimitrije Marković 57/2019

Matematički fakultet u Beogradu  
20.6.2024

# Uvod

U današnjem digitalnom dobu, sposobnost automatskog prepoznavanja i identifikovanja lica postaje sve važnija za različite primene, uključujući bezbednost, marketing, zabavu i personalizovane usluge.

Naša studija se fokusira na prepoznavanje lica poznatih ličnosti, što je posebno izazovan zadatak zbog raznih faktora kao što su različiti izrazi lica, promene u frizuri, šminka, kao i različiti osvetljenja i uglovi snimanja.

# Ciljevi projekta

Treniranje neuronske mreže za prepoznavanje poznatih ličnosti sa slika.

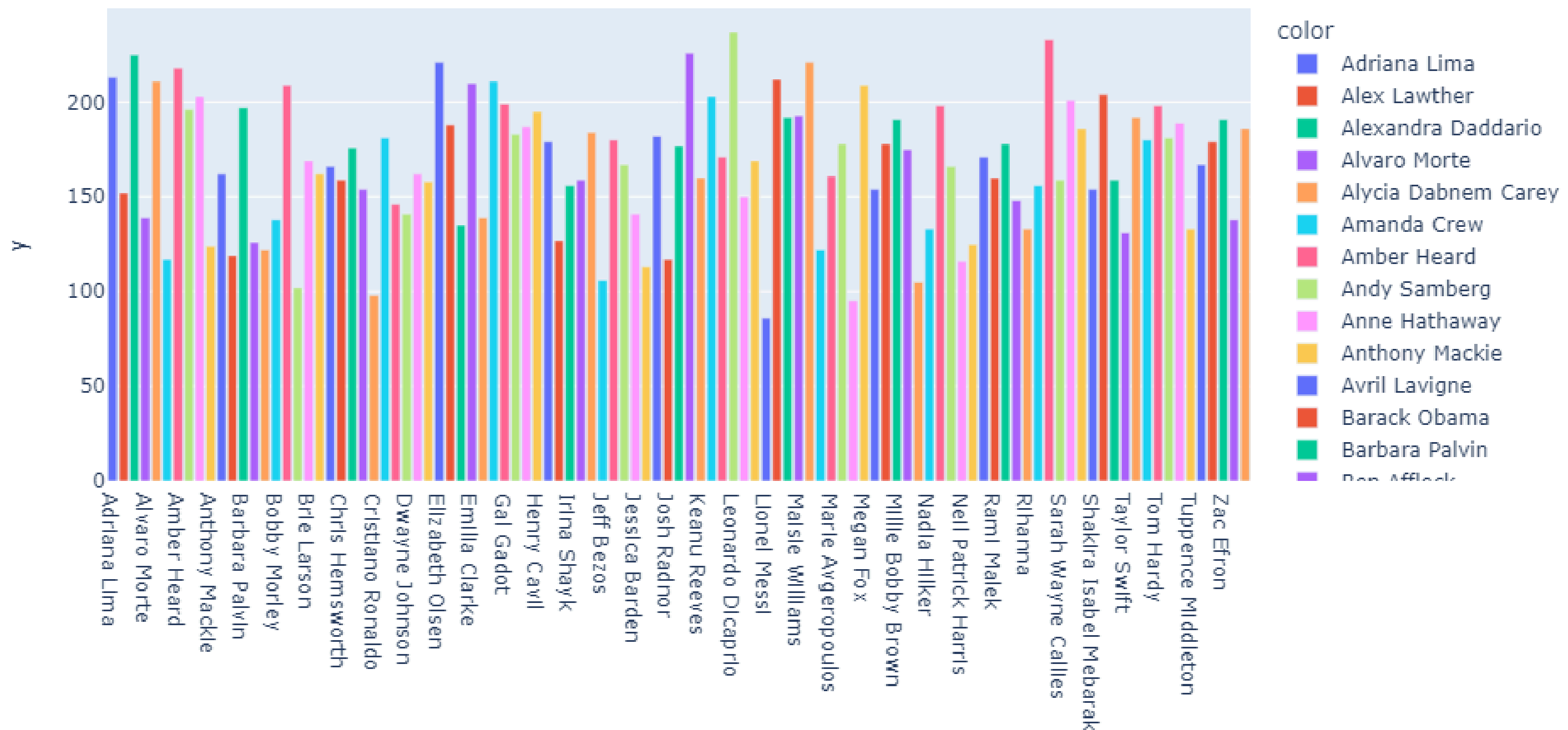
Primena tehnika preprocesiranja podataka, optimizacije modela i evaluacije performansi kako bismo postigli visoku tačnost prepoznavanja.

Upoređivanje rezultata kako bismo utvrdili koji model (VGG ili ResNet) se može bolje fine-tunovati za ovaj zadatak.

# Skup podataka

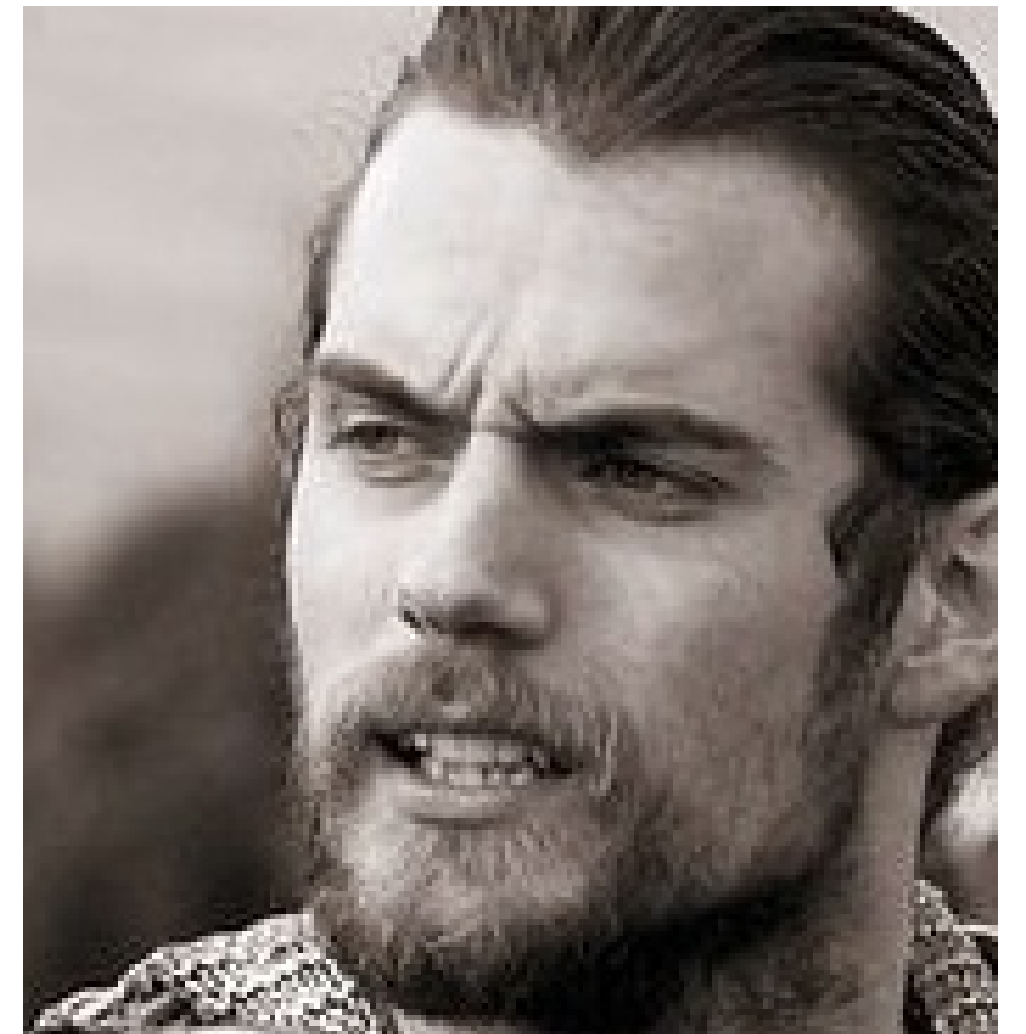
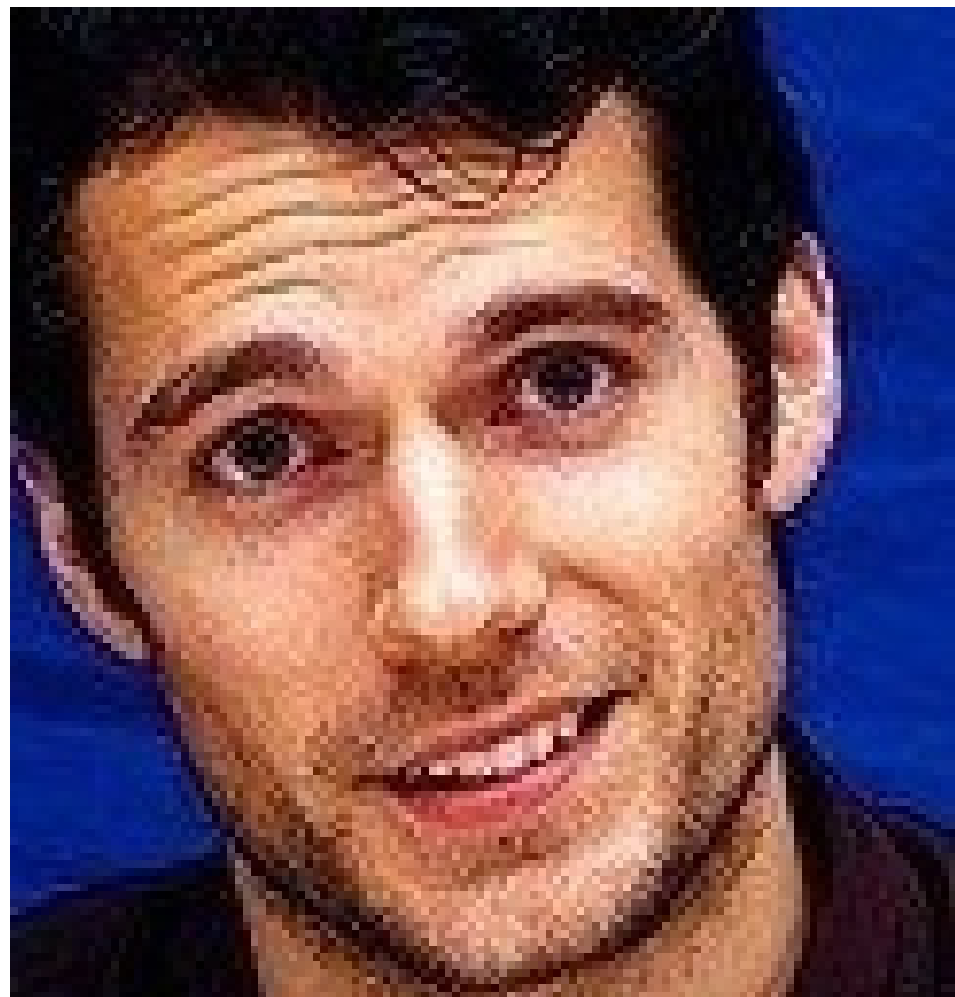
U skupu podataka imamo slike 105 poznatih ličnosti, za svaku ličnost imamo od 86 do 237 slika. Ukupno 17354 slike

Distribution of number of images per person



# Podela skupa podataka

U našem projektu prepoznavanja lica glumaca, odlučili smo da podelimo skup podataka na trening(65%), validacioni(15%) i deo za test(20%) bez korišćenja unakrsne validacije. S obzirom na to da imamo veliki skup podataka, dovoljan je slučajan uzorak za testiranje i validaciju modela. Ova podela nam omogućava da efikasno treniramo i procenjujemo naš model bez potrebe za dodatnim računskim opterećenjem koje bi donela unakrsna validacija.



# Preprocesiranje podataka

U procesu preprocesiranja skupa podataka za naš projekat prepoznavanja glumaca, primenili smo tri ključne transformacije:

- **Skaliranje:** Sve slike smo skalirali na istu dimenziju od 224x224 piksela. Ovo osigurava konzistentnost u ulaznim podacima za naš model.
- **Konverzija u tenzore:** Svaku sliku smo konvertovali iz formata PIL Image u tenzor, što je osnovni format podataka za obradu u PyTorch-u.
- **Normalizacija:** Normalizovali smo tenzore slika koristeći srednje vrednosti [0.485, 0.456, 0.406] i standardne devijacije [0.229, 0.224, 0.225]. Ova transformacija pomaže stabilizaciji treninga modela tako što smanjuje varijaciju u podacima i omogućava bržu konvergenciju.

Nismo koristili augmentaciju podataka u našem projektu jer smo imali veliki skup podataka, koji je bio dovoljno raznovrstan i obiman da omogući treniranje modela bez potrebe za dodatnim tehnikama augmentacije.

# Metodologije

Izabrali smo fine-tuning pre-treniranih modela (VGG i ResNet) zbog relativno malog skupa podataka.

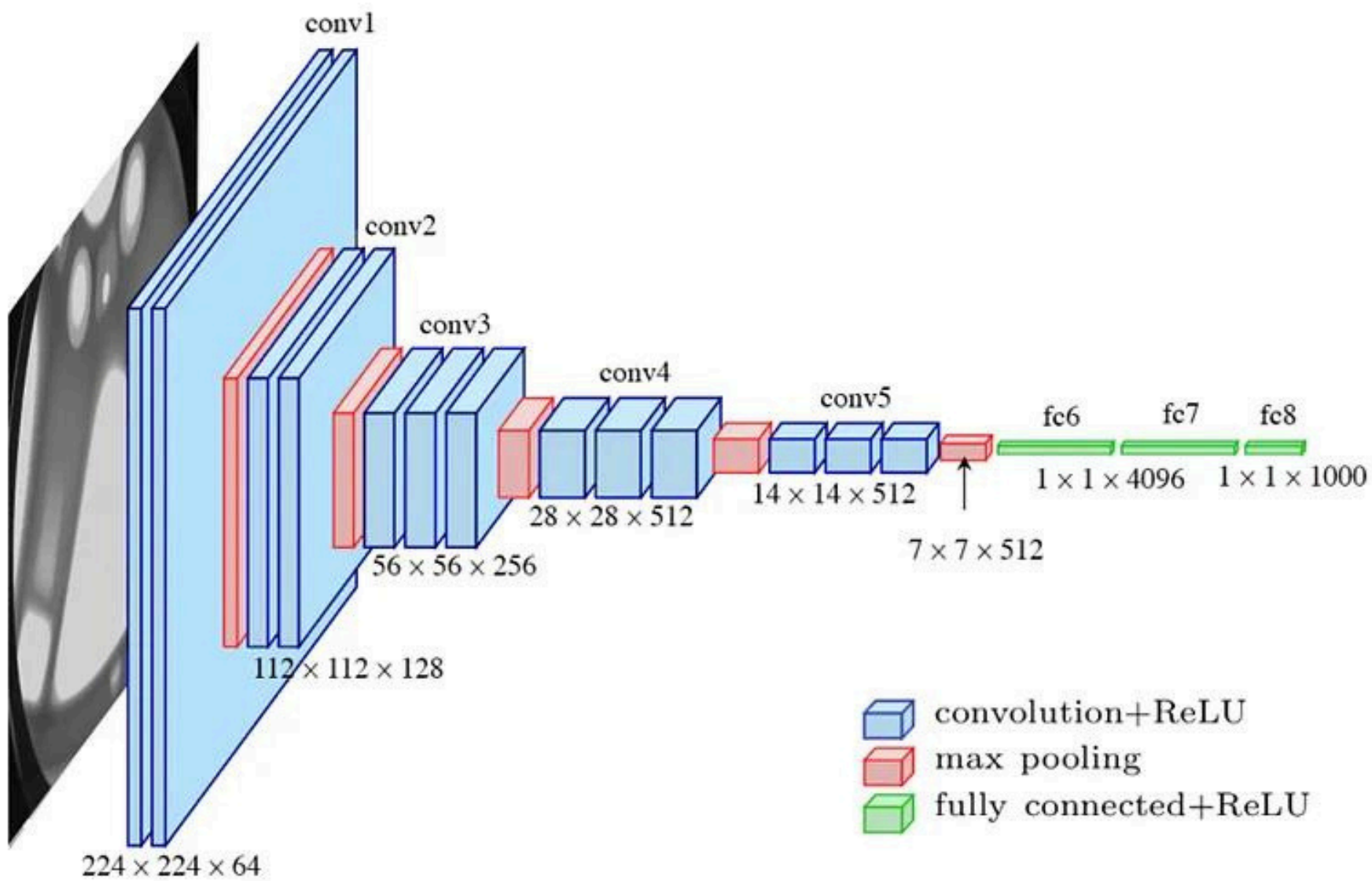
Fine-tuning omogućava korišćenje već naučenih karakteristika iz velikih datasetova, kao što je ImageNet, čime se smanjuje potreba za ogromnim količinama podataka i računske snage

Prednosti fine-tunovanja uključuju brže treniranje, bolje performanse na malim skupovima podataka i mogućnost prilagođavanja modela specifičnom zadatku prepoznavanja glumaca.



# VGG

VGG je popularna arhitektura konvolucionih neuronskih mreža (CNN) koju su razvili istraživači iz Visual Geometry Group na Univerzitetu u Oksfordu. VGG modeli, kao što su VGG16 i VGG19, poznati su po svojoj jednostavnoj i uniformnoj strukturi, koja koristi male 3x3 filtere u svim konvolucionim slojevima. Ovi modeli su postigli vrhunske rezultate na mnogim zadacima prepoznavanja slika, uključujući i takmičenje ImageNet. Prednost VGG modela je u njihovoj dubini i jednostavnosti, što ih čini pogodnim za fine-tuning i primenu na specifične zadatke sa relativno malim skupovima podataka.

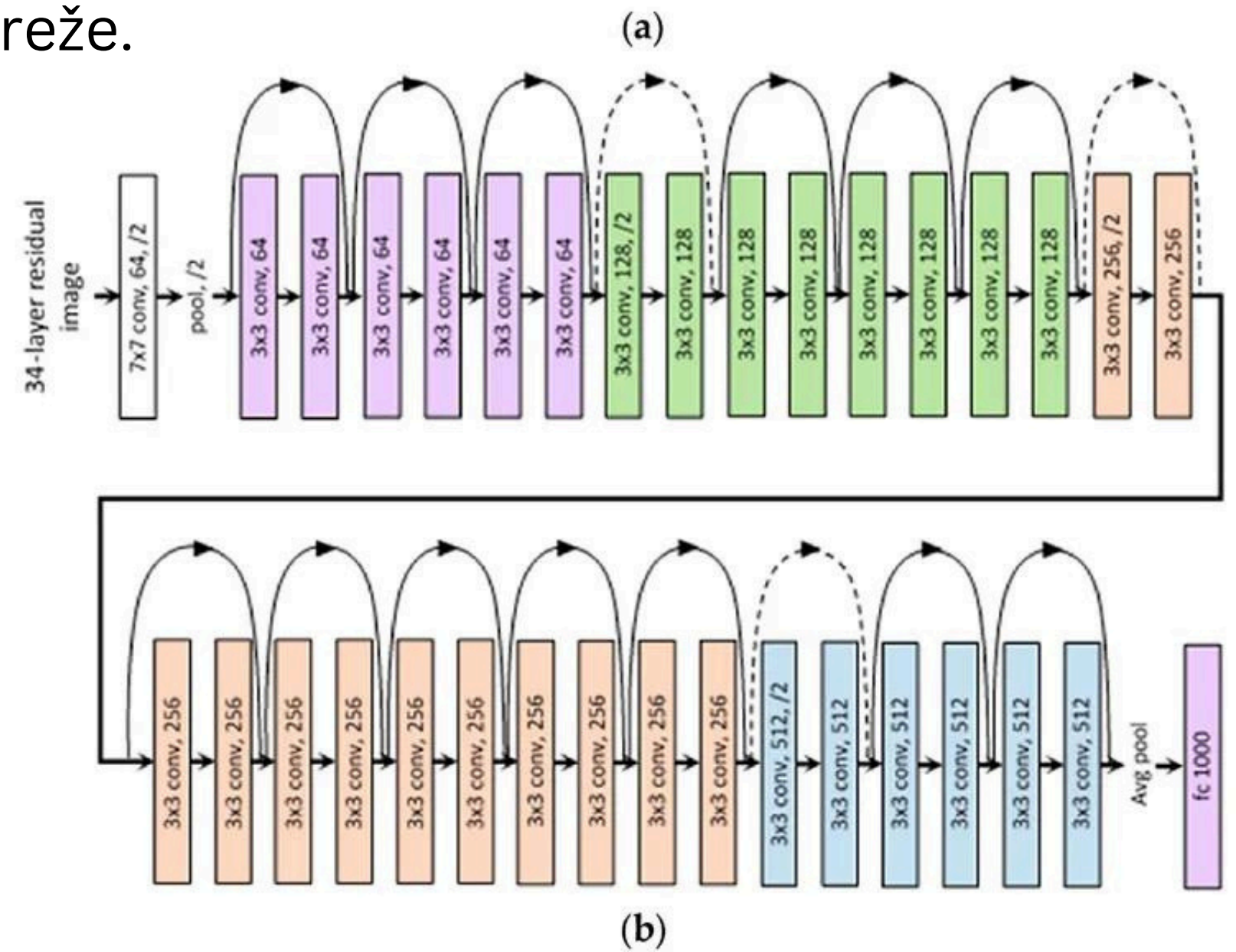


	Layer	Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	224 x 224 x 3	-	-	-
1	2 X Convolution	64	224 x 224 x 64	3x3	1	relu
	Max Pooling	64	112 x 112 x 64	3x3	2	relu
3	2 X Convolution	128	112 x 112 x 128	3x3	1	relu
	Max Pooling	128	56 x 56 x 128	3x3	2	relu
5	2 X Convolution	256	56 x 56 x 256	3x3	1	relu
	Max Pooling	256	28 x 28 x 256	3x3	2	relu
7	3 X Convolution	512	28 x 28 x 512	3x3	1	relu
	Max Pooling	512	14 x 14 x 512	3x3	2	relu
10	3 X Convolution	512	14 x 14 x 512	3x3	1	relu
	Max Pooling	512	7 x 7 x 512	3x3	2	relu
13	FC	-	25088	-	-	relu
14	FC	-	4096	-	-	relu
15	FC	-	4096	-	-	relu
Output	FC	-	1000	-	-	Softmax



# Resnet

ResNet (Residual Network) je napredna arhitektura konvolucionih neuronskih mreža (CNN) koju su razvili istraživači iz Microsoft Research-a. ResNet modeli, kao što su ResNet50 i ResNet101, uveli su koncept rezidualnih blokova koji omogućavaju treniranje vrlo dubokih mreža bez problema degradacije performansi. Rezidualni blokovi koriste shortcut konekcije koje preskaču jedan ili više slojeva, čime olakšavaju optimizaciju i omogućavaju efikasnije učenje. ResNet modeli su postigli izvanredne rezultate na mnogim zadacima, uključujući i takmičenje ImageNet, i postali su standard u dubokom učenju zbog svoje sposobnosti da efikasno treniraju veoma duboke mreže.



# Eksperimentalni rezultati

Modeli bez fine-tuninga

	<b>Resnet</b>	<b>VGG</b>
Accuracy	0.00987	0.01177
Loss	4.7017	4.6935

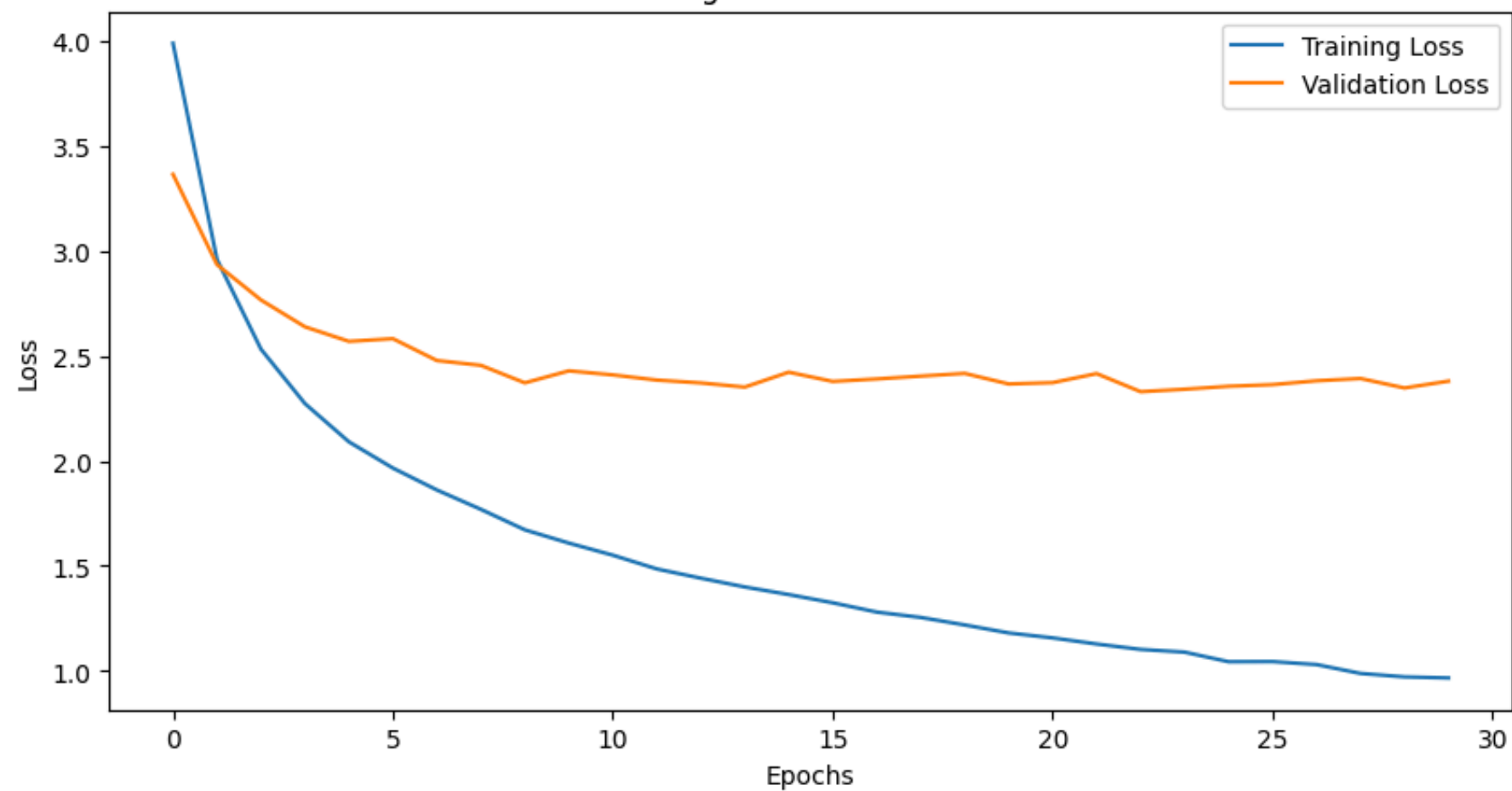
Vidimo da su originalni modeli bez treniranja, ocekivano, dali baš loše rezultate

Fine-tunovanje samo poslednjih potpuno povezanih slojeva

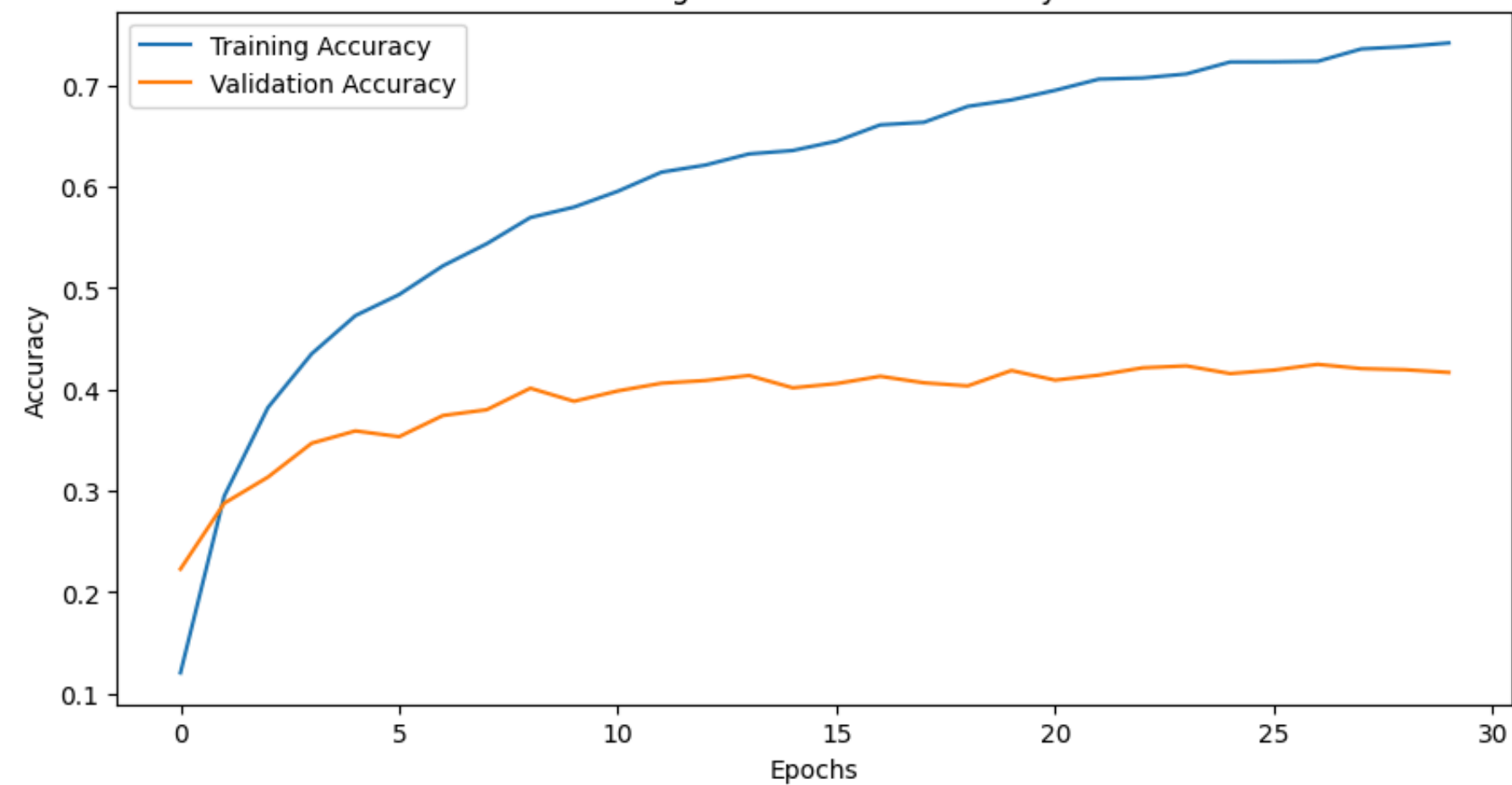
	Resnet	VGG
Accuracy	0.4169	0.2809
Loss	2.3798	3.5692

# RESNET plotovi

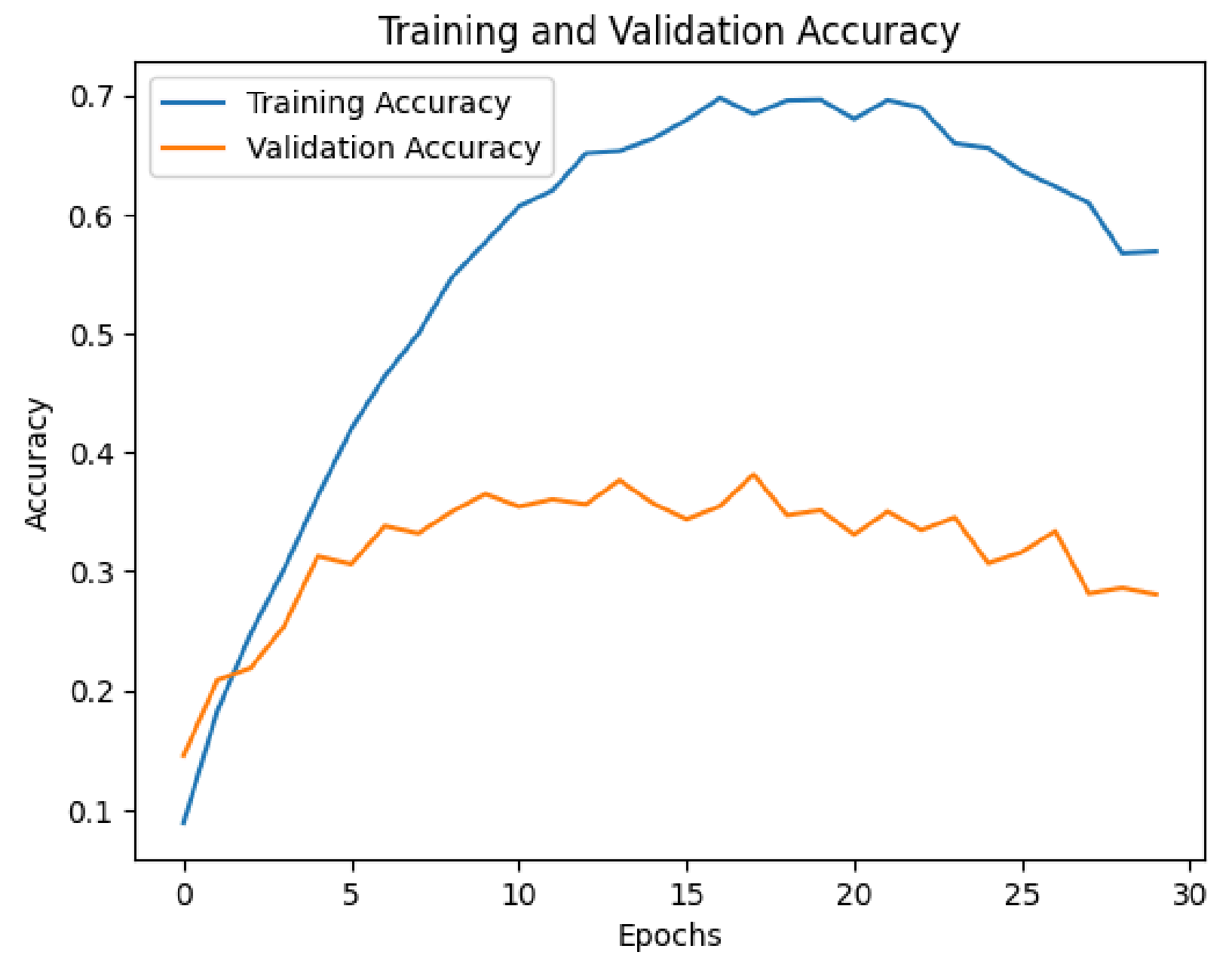
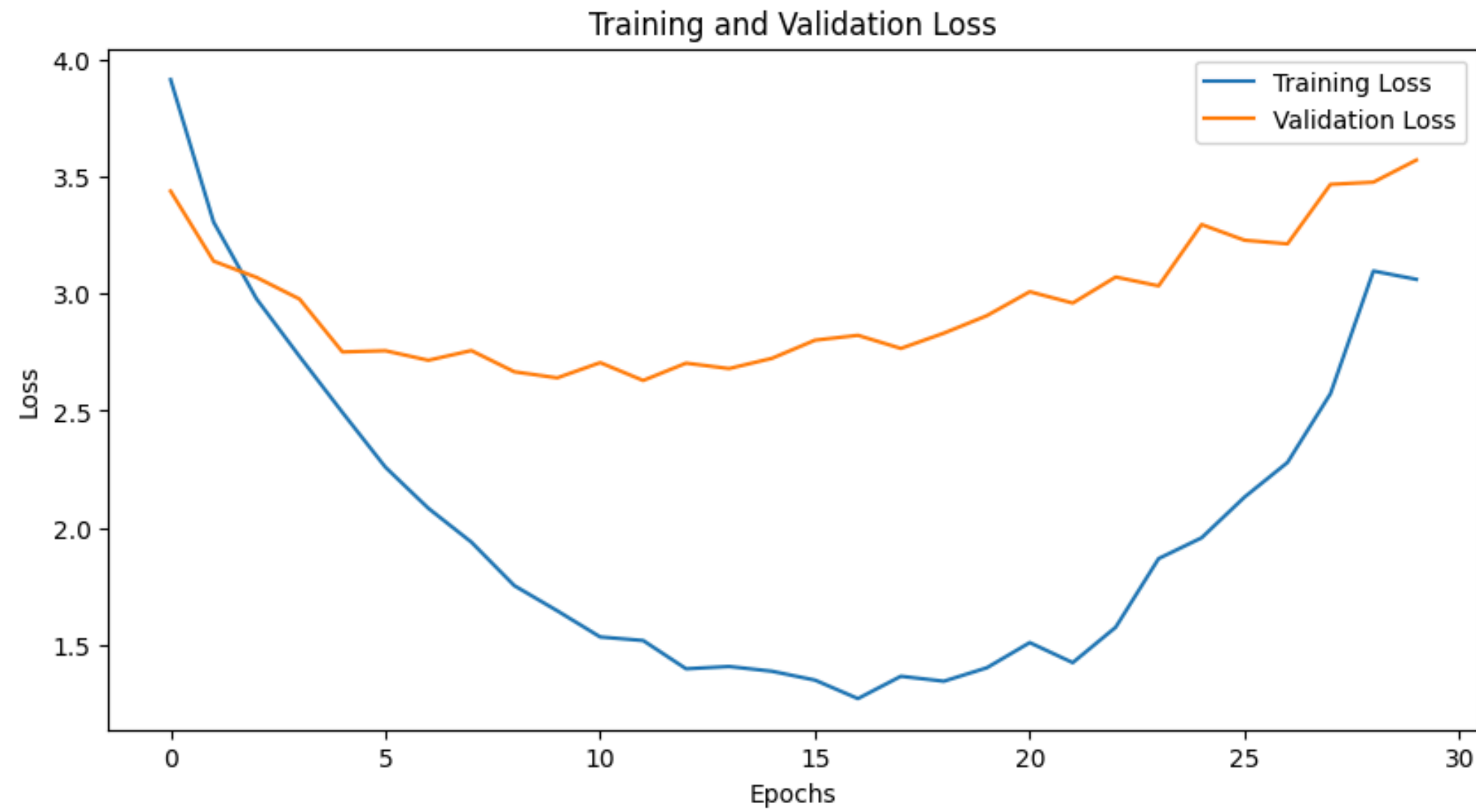
Training and Validation Loss



Training and Validation Accuracy



# VGG plotovi



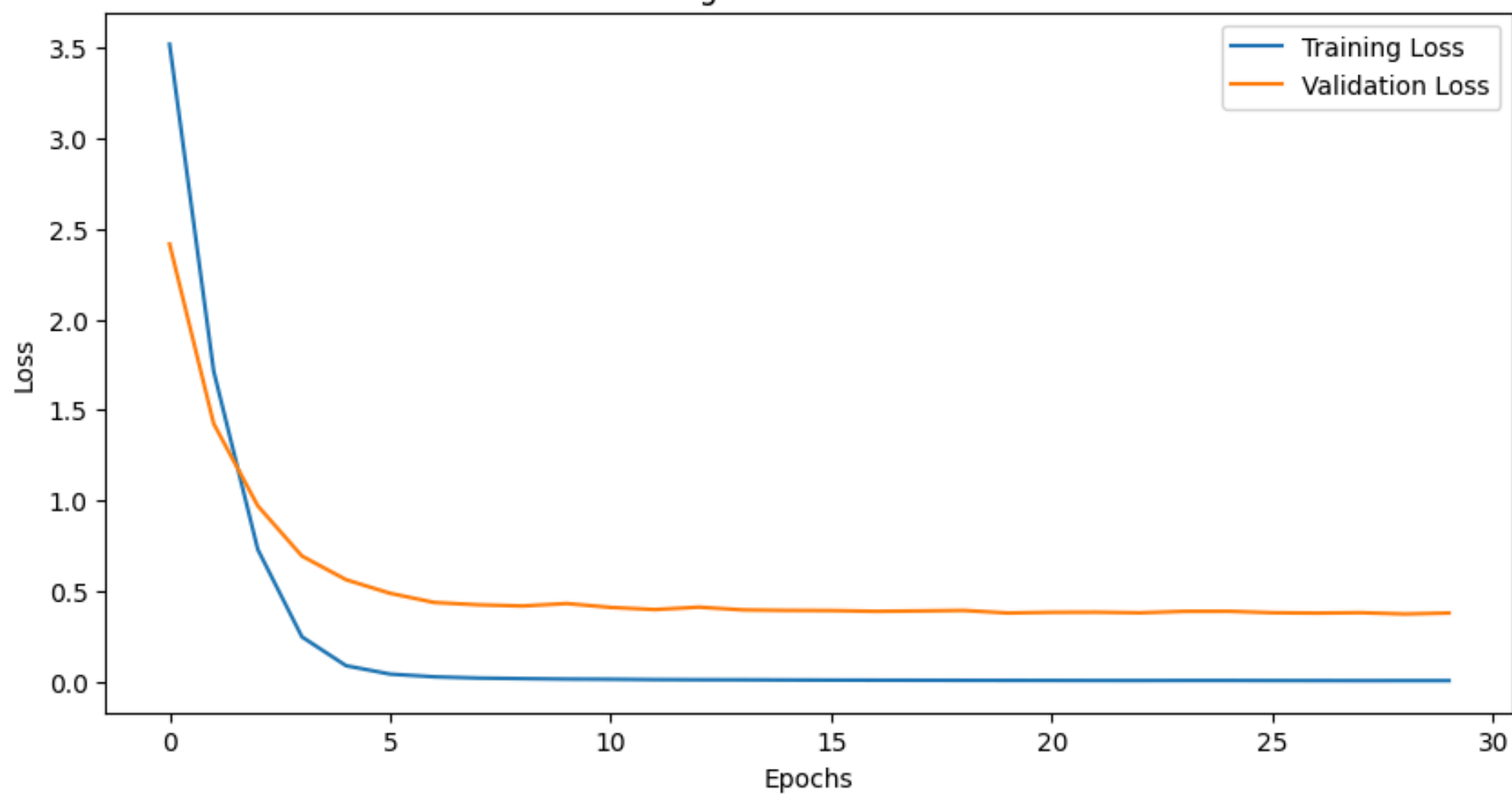
## Fine-tunovanje samo konvolutivnih slojeva

	<b>Resnet</b>	<b>VGG</b>
Accuracy	0.9115	0.8132
Loss	0.3769	0.9116

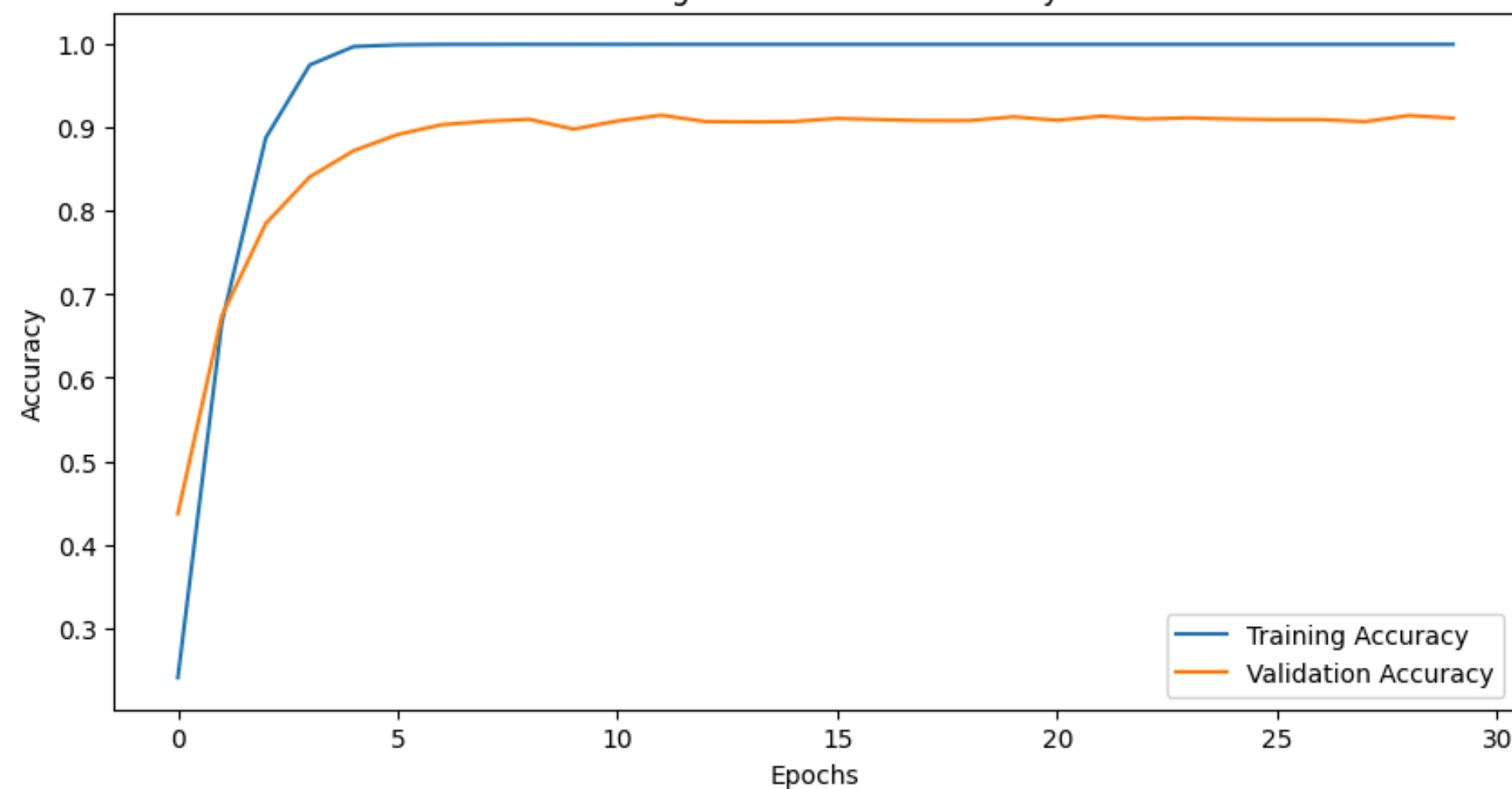


# RESNET plotovi

Training and Validation Loss

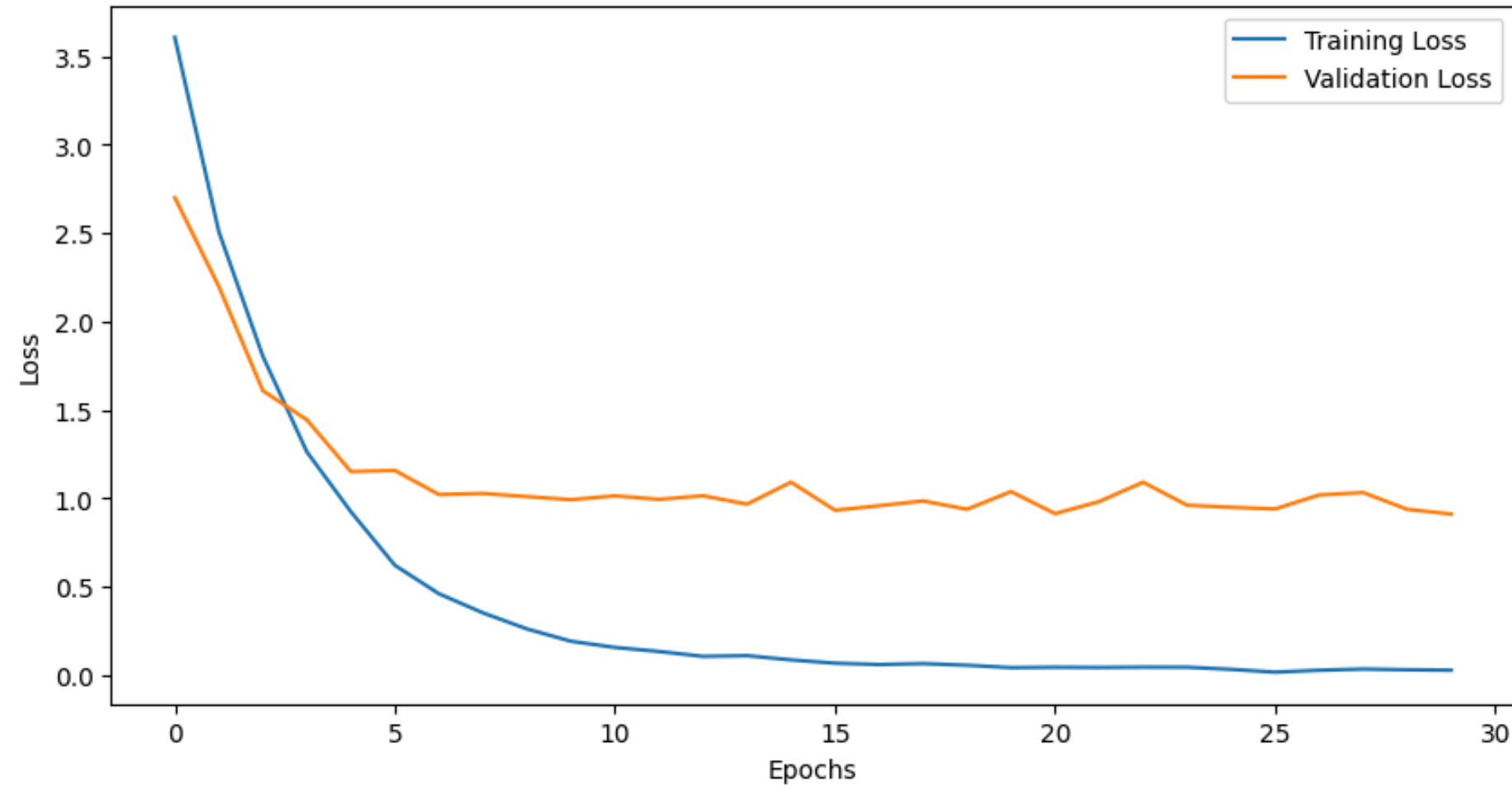


Training and Validation Accuracy

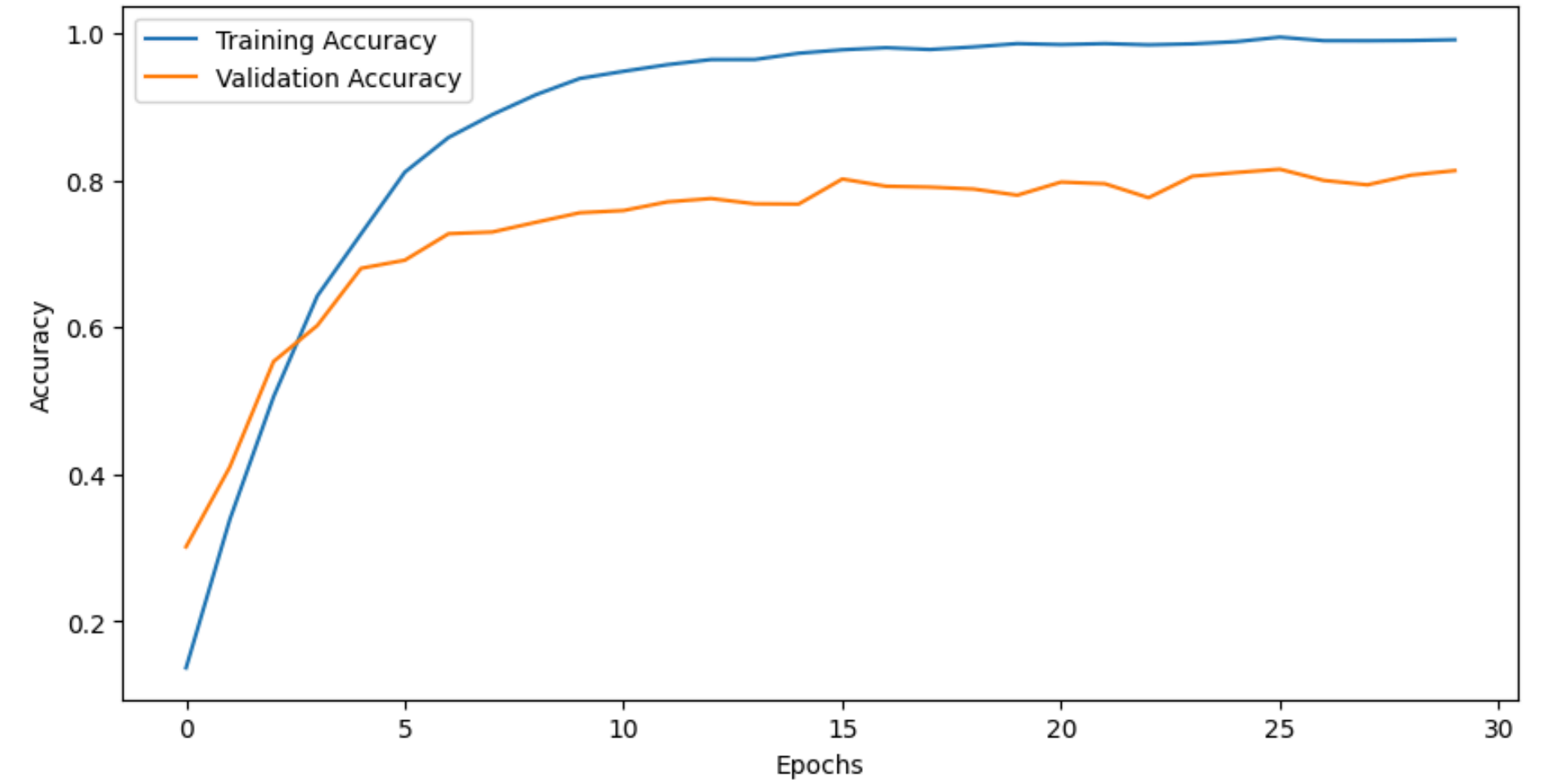


# VGG plotovi

Training and Validation Loss VGG



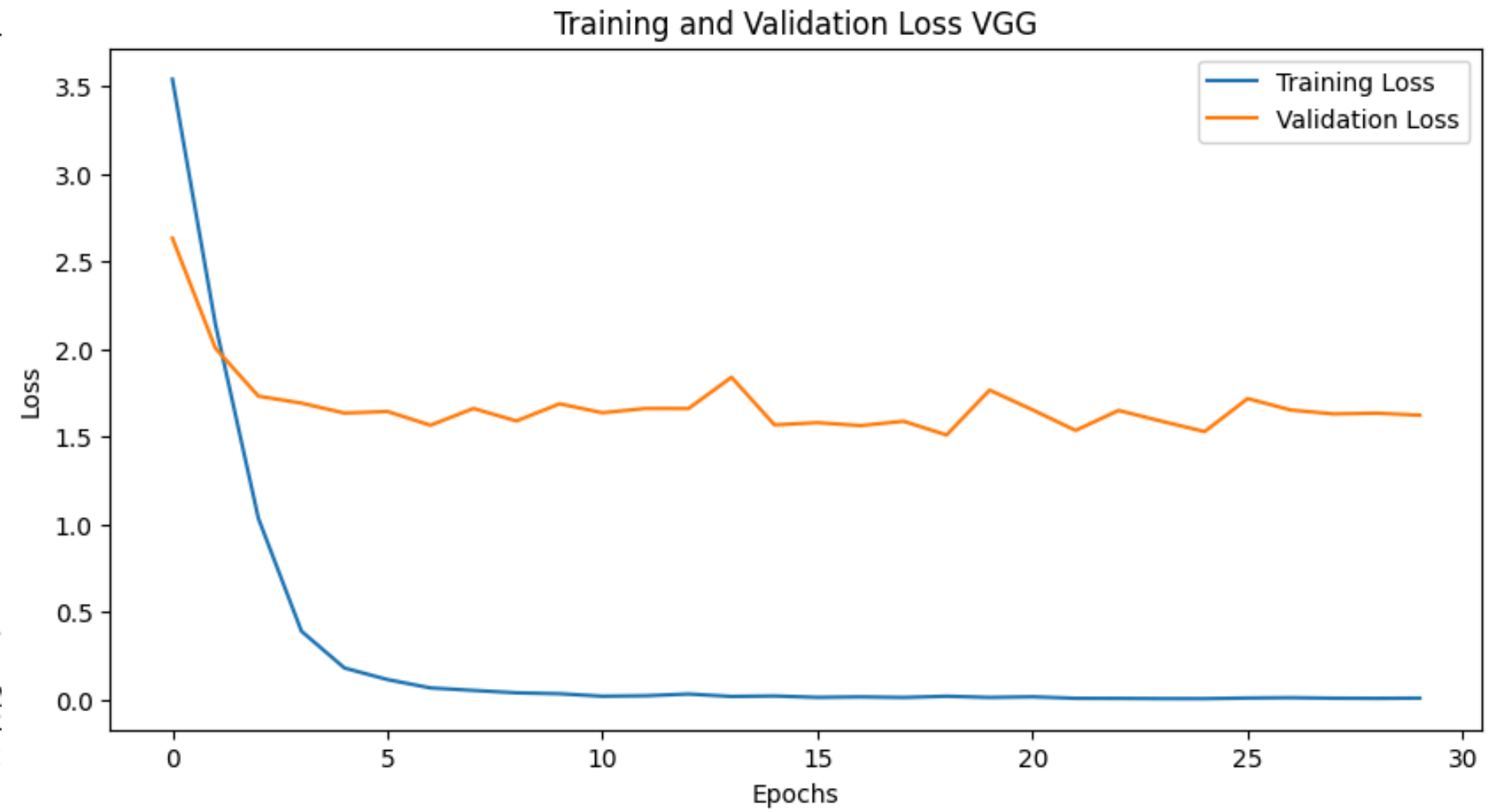
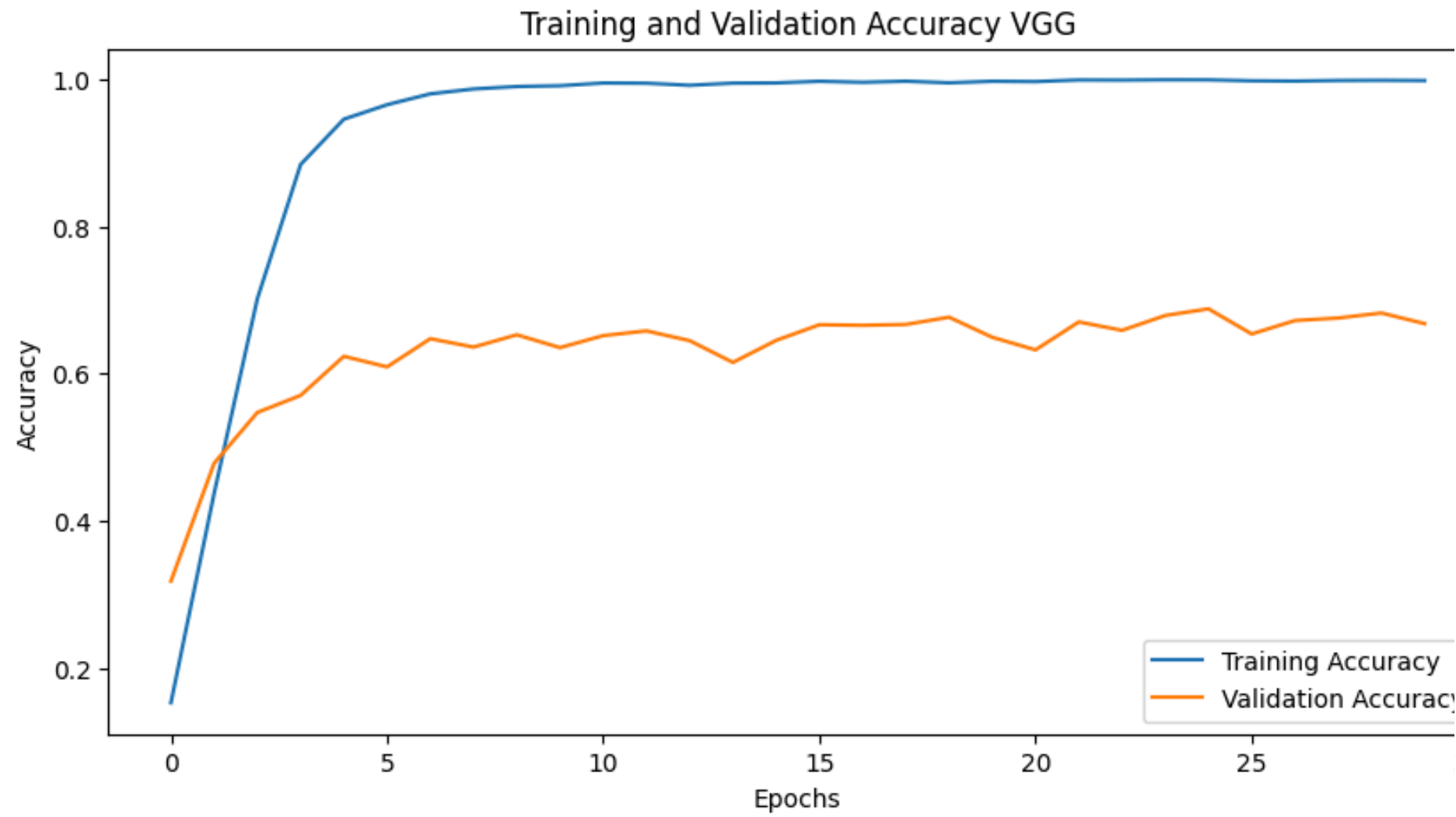
Training and Validation Accuracy VGG



## Fine-tunovanje i konvolutivnih slojeva i fully connected slojeva

	<b>VGG</b>
Accuracy	0.6678
Loss	1.6218

# VGG plotovi



# Optimizacija

Nakon svih različitih testiranja dobili smo rezultate da su najprecizniji modeli koji se dobijaju fine-tunovanjem samo konvolucijskih slojeva.

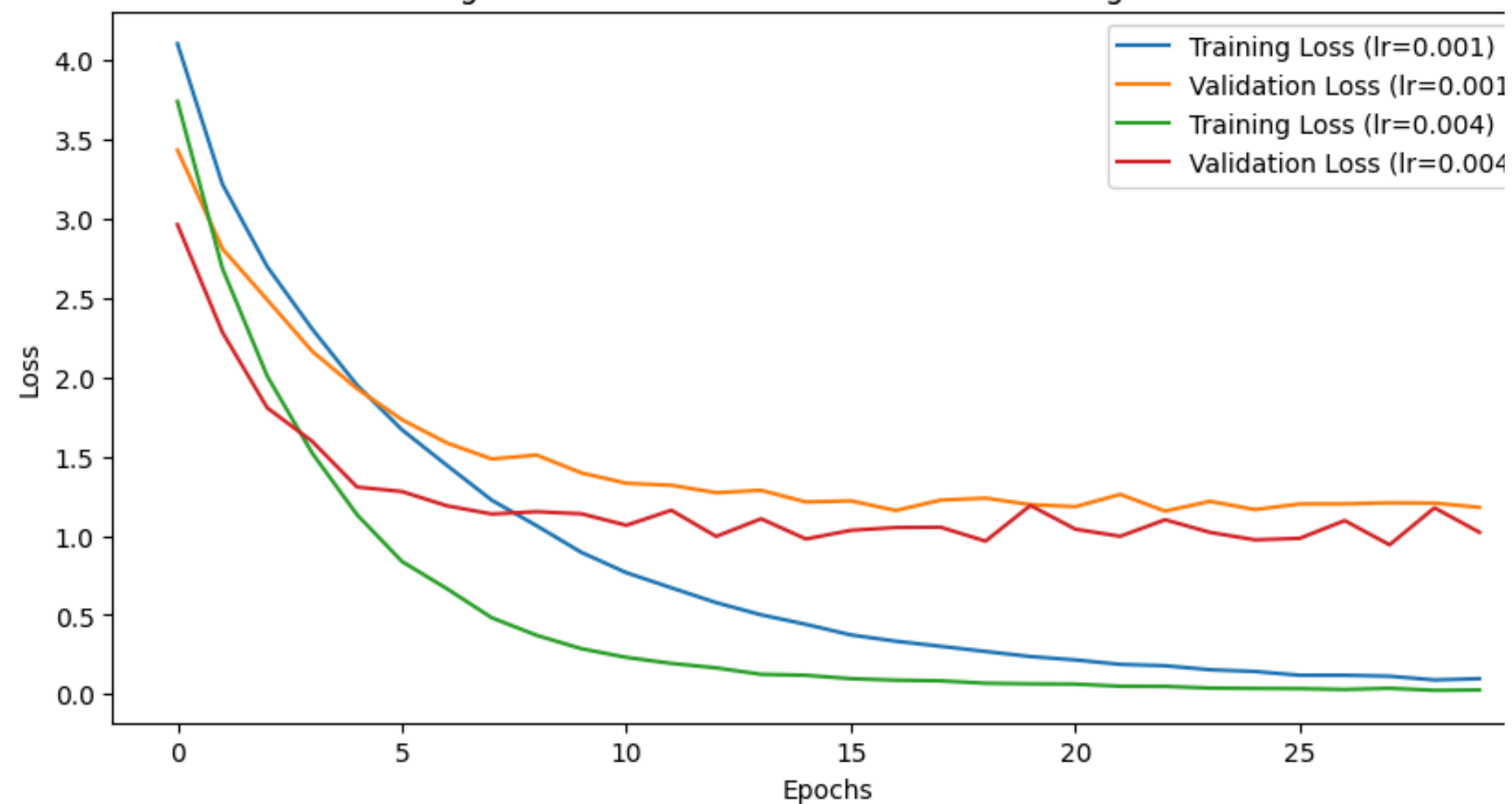
	Resnet	VGG
Accuracy	0.9115	0.8132
Loss	0.3769	0.9116

Dodatno, najbolje modele smo optimizovali podešavanjem hiperparametra brzine učenja, za različitu vrednost learning rate-a (0.001, 0.004, 0.008).

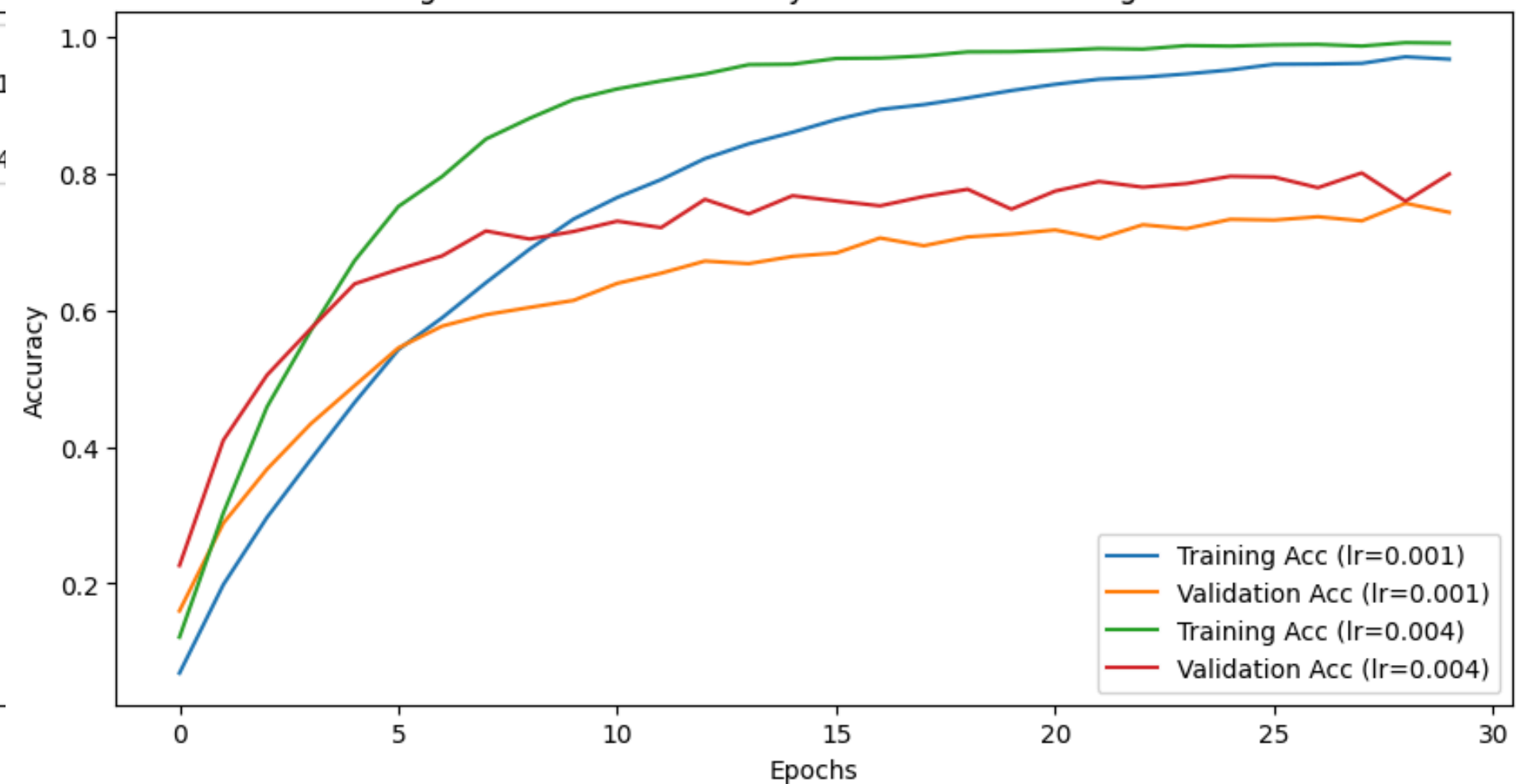
VGG	lr=0.001	lr=0.004	lr=0.008
accuracy	0.7437	0.7999	0.8132
loss	1.1789	1.0220	0.9116
RESNET	lr=0.001	lr=0.004	lr=0.008
accuracy	0.8394	0.8785	0.9115
loss	0.8326	0.5222	0.3769



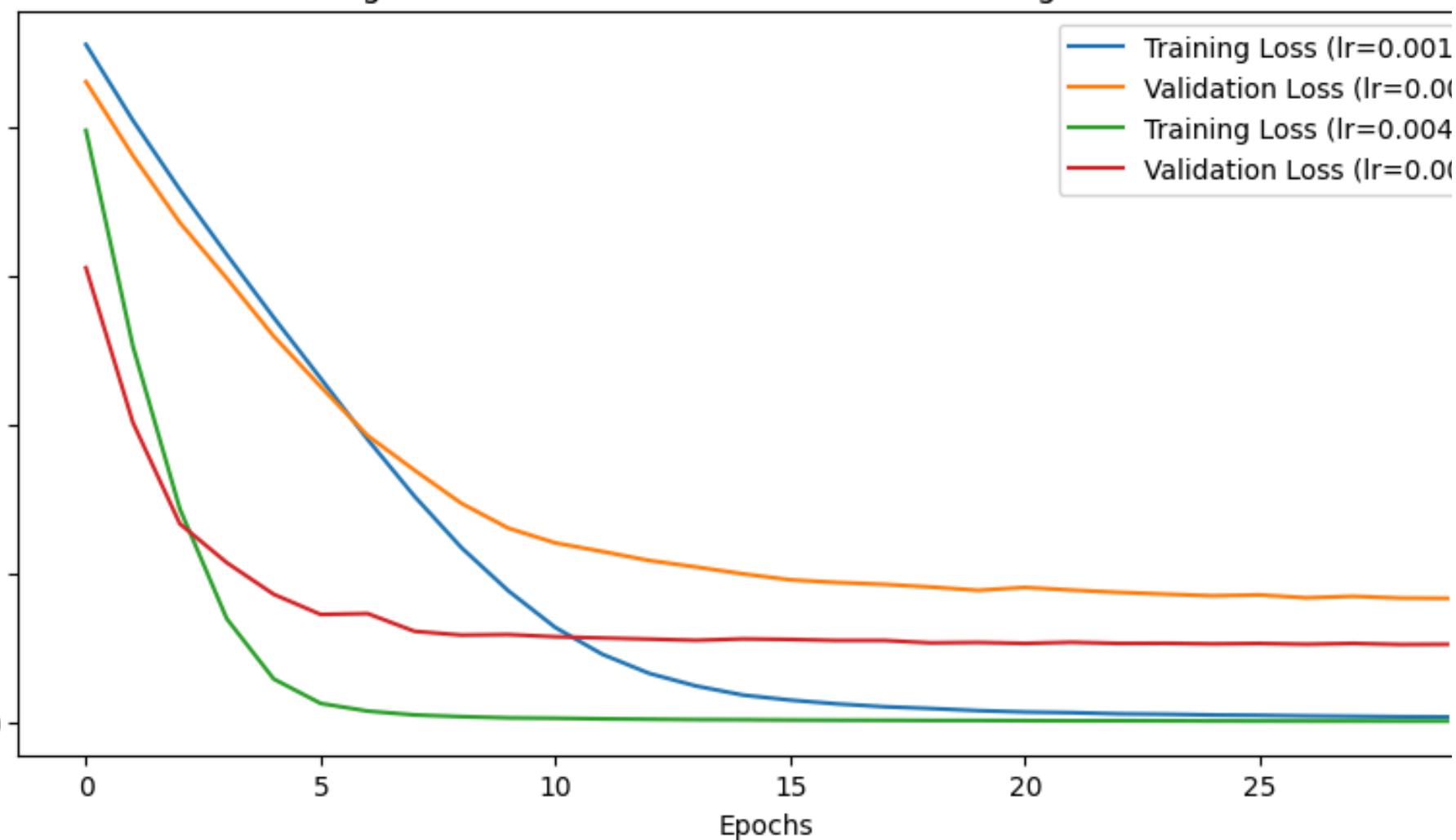
Training and Validation Loss for Different Learning Rates VGG



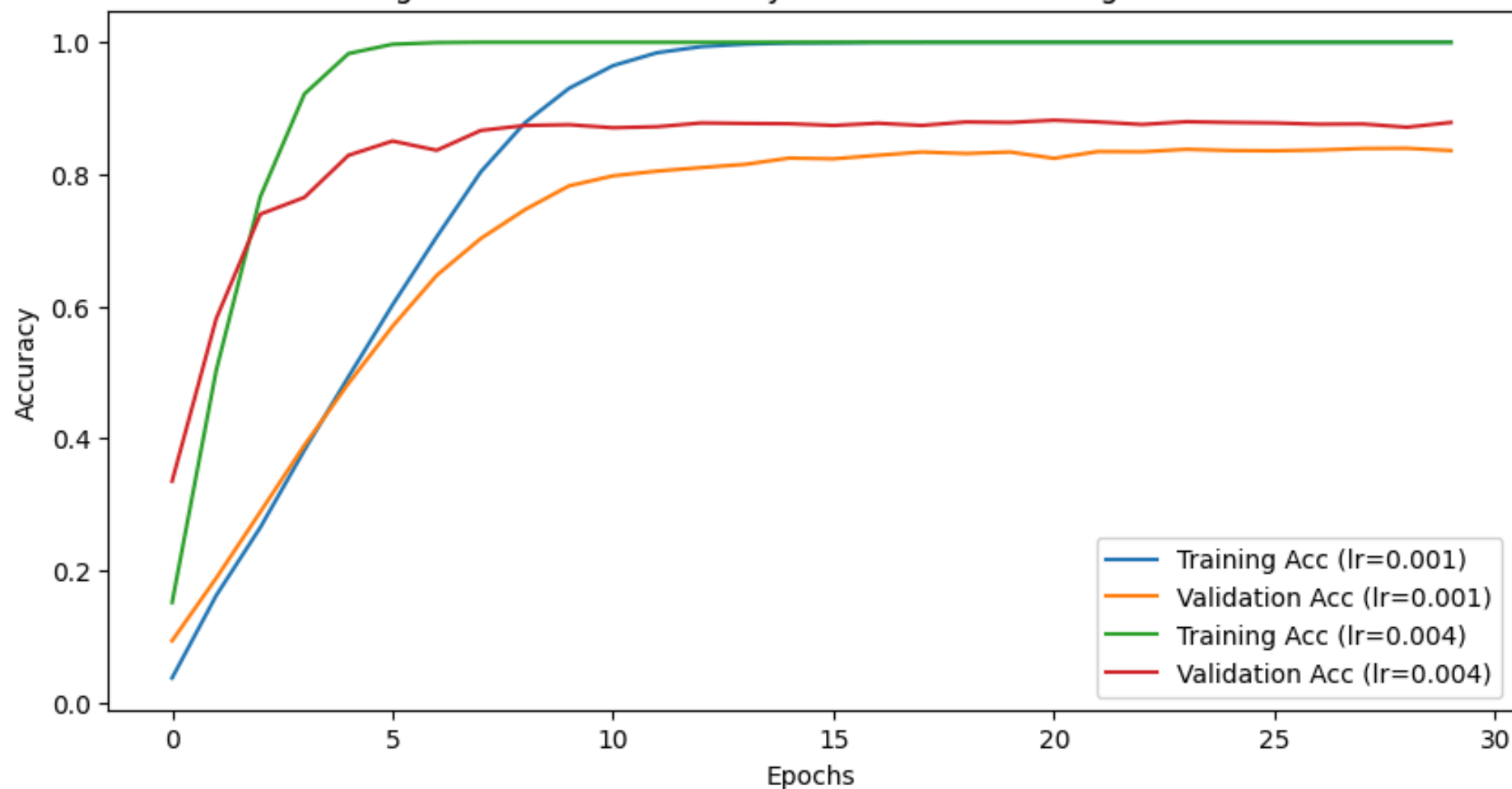
Training and Validation Accuracy for Different Learning Rates VGG



Training and Validation Loss for Different Learning Rates Resnet



Training and Validation Accuracy for Different Learning Rates Resnet



# Evaluacija

Od svih modela koje smo trenirali, izabrali smo najbolji model na osnovu performansi na validacionom skupu. Ovaj model smo potom evaluirali na test skupu kako bismo proverili da li je došlo do preprilagođavanja (overfittinga) i da li model generalizuje dobro na neviđene podatke.

Rezultati evaluacije na test skupu pokazali su se konzistentnim sa rezultatima na validacionom skupu, što ukazuje da naš model nije preprilagođen i da dobro generalizuje na nove podatke. Ovo potvrđuje da je naš pristup treniranju i fine-tuningu modela bio efikasan, te da su naši modeli spremni za primenu u stvarnim scenarijima prepoznavanja glumaca sa slika.

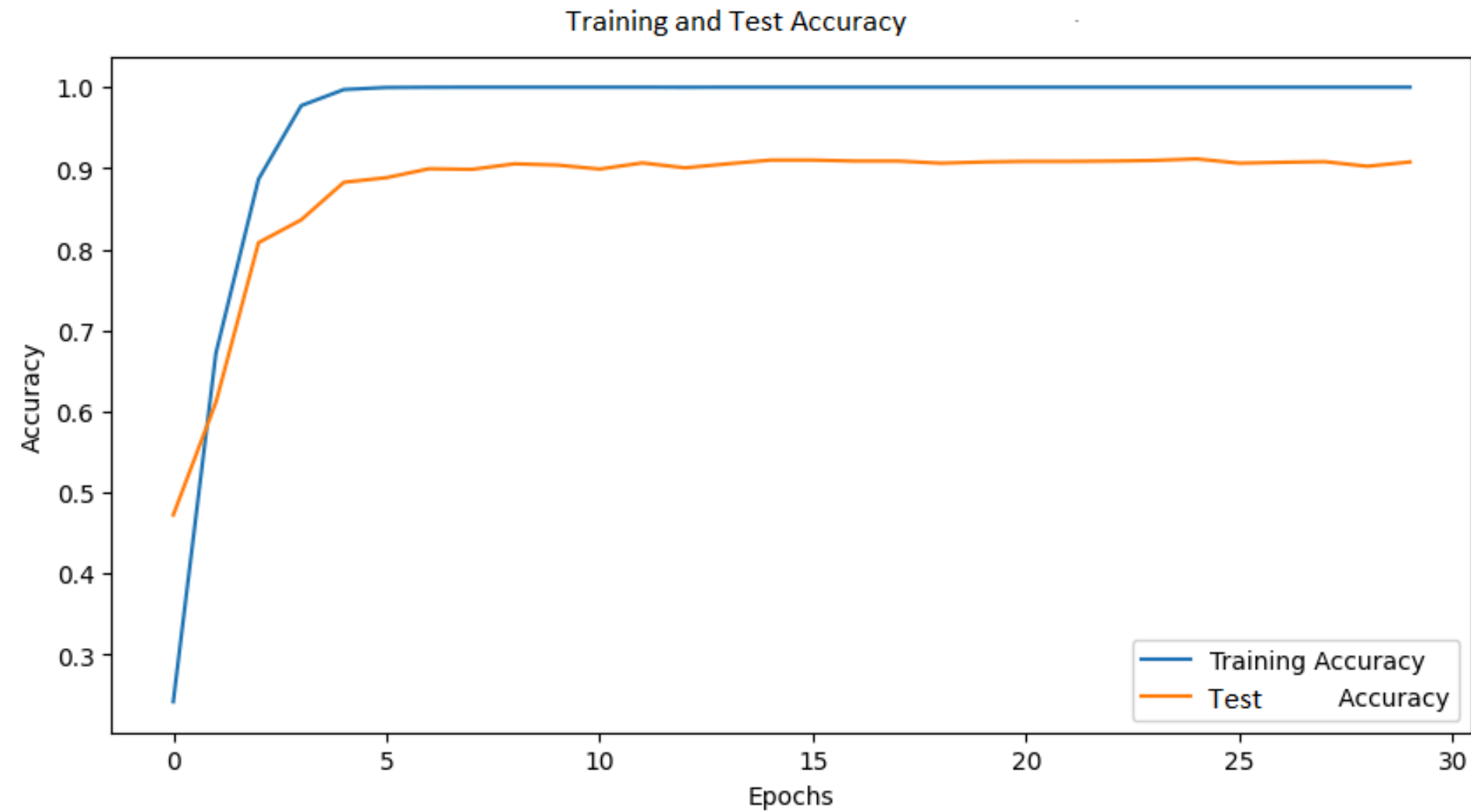
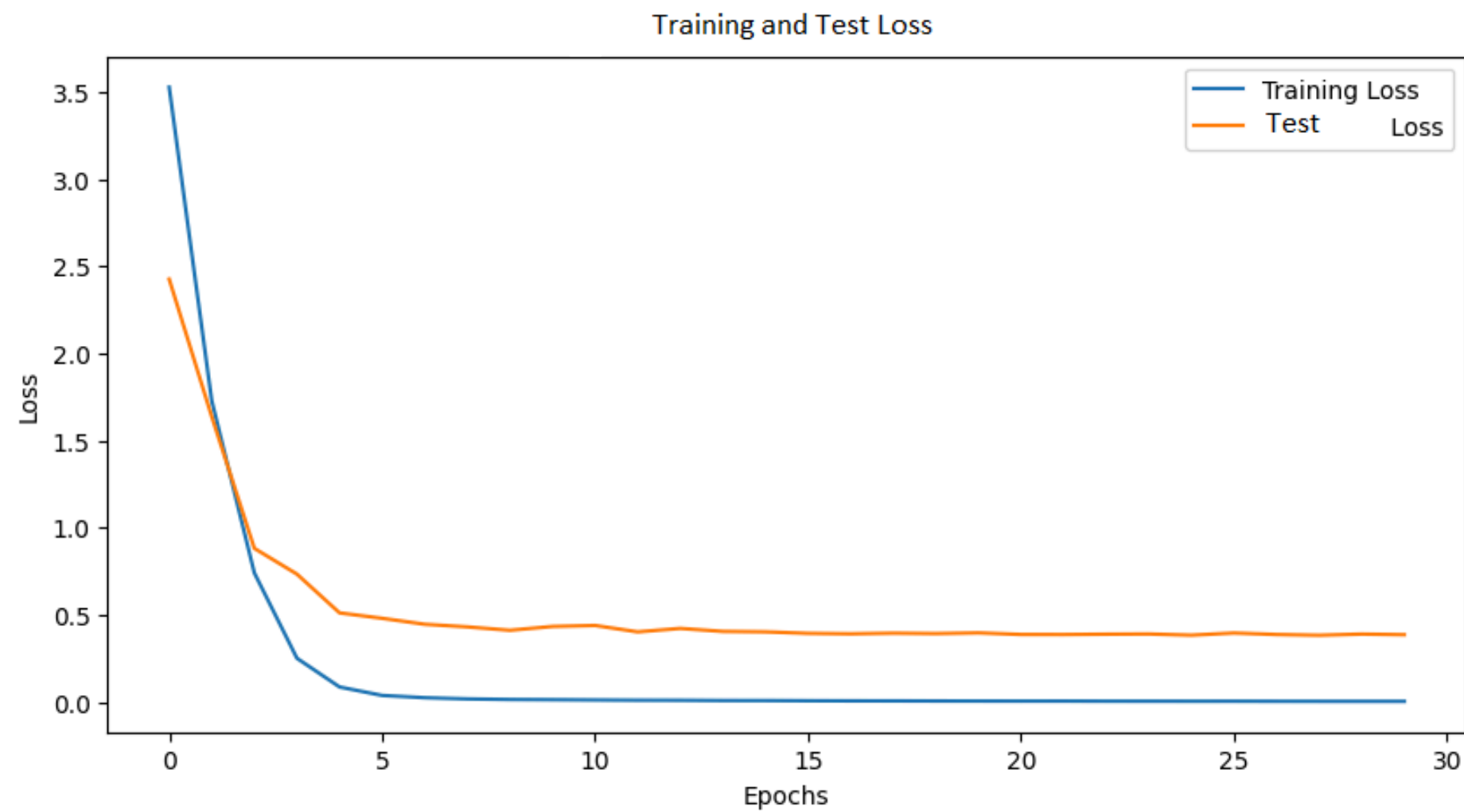
## ***Test metrics:***

Accuracy: 0.8940

Precision: 0.8980

Recall: 0.8940

F1 Score: 0.8939



# Zaključak

Naš projekat prepoznavanja glumaca sa slika pokazao je da se najbolji rezultati postižu kada treniramo samo konvolucijske blokove VGG i ResNet arhitektura. Iako je intuitivno očekivati da će treniranje samo potpuno povezanih (fully connected) slojeva biti dovoljno za prilagođavanje specifičnom zadatku, naši eksperimenti su pokazali da uključivanje konvolucijskih slojeva omogućava modelima da nauče dublje i složenije karakteristike potrebne za prepoznavanje lica.

Neki od mogućih razloga su:

- Efikasnije Korišćenje Podataka
- Regularizacija i Generalizacija
- Dublje Razumevanje Slika

Pri upoređivanju VGG i ResNet arhitektura, primetili smo da ResNet postiže nešto bolje rezultate. To se može objasniti činjenicom da ResNet koristi rezidualne blokove koji omogućavaju efikasnije treniranje dubljih mreža bez problema degradacije performansi. Rezidualne konekcije olakšavaju propagaciju gradijenata kroz mrežu, što omogućava modelu da nauči bogatije reprezentacije i bolje generalizuje na nove podatke.

Sveukupno, naš rad pokazuje da fine-tuning pre-treniranih modela, posebno kada uključuje treniranje konvolucijskih slojeva (u našem slučaju), može pružiti odlične rezultate čak i sa relativno malim skupom podataka. ResNet se istakao kao superiorna arhitektura u našem zadatku prepoznavanja glumaca, pružajući robusne i pouzdane performanse.



**Hvala na pažnji!**