

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

I ПОДАЦИ О КОМИСИЈИ		
1. Датум и орган који је именовео комисију:		
<b>Решење број 012-199/31-2015 од 31.03.2022.</b> , на основу одлуке Наставно Научног већа факултета, декан Факултета техничких наука, проф. др Срђан Колаковић		
2. Састав комисије у складу са <i>Правилима докторских студија Универзитета у Новом Саду</i> :		
1. Др Иван Мезеи	Ванредни професор	Електроника, 20.12.2017.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Универзитет у Новом Саду	Председник	
установа у којој је запослен-а		функција у комисији
2. Др Дејан Вукобратовић	Редовни професор	Телекомуникације и обрада сигнала, 01.04.2019.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Универзитет у Новом Саду	Члан	
установа у којој је запослен-а		функција у комисији
3. Др Татјана Николић	Редовни професор	Електроника, 27.10.2021.
презиме и име	звање	ужа научна област и датум избора
Електронски факултет, Универзитет у Нишу	Члан	
установа у којој је запослен-а		функција у комисији
4. Др Вук Врањковић	Ванредни професор	Електроника, 24.09.2020.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Универзитет у Новом Саду	Члан	
установа у којој је запослен-а		функција у комисији
5. Др Растислав Струхарик	Редовни професор	Електроника, 17.07.2020.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Универзитет у Новом Саду	Ментор	
установа у којој је запослен-а		функција у комисији

<b>II ПОДАЦИ О КАНДИДАТУ</b>
<p>1. Име, име једног родитеља, презиме: <b>Дамјан, Мирјана, Ракановић</b></p> <p>2. Датум рођења, општина, држава: <b>17.03.1991., Нови Сад, Република Србија</b></p> <p>3. Назив факултета, назив претходно завршеног нивоа студија и стечени стручни/академски назив: <b>Факултет техничких наука, Универзитет у Новом Саду, Енергетика, електроника и телекомуникације, Мастер инжењер електротехнике и рачунарства</b></p> <p>4. Година уписа на докторске студије и назив студијског програма докторских студија: <b>2015., Енергетика, електроника и телекомуникације</b></p>
<b>III НАСЛОВ ДОКТОРСKE ДИСЕРТАЦИЈЕ:</b>
<b>Хардверска акцелерација конволуционих неуронских мрежа у ембедед системима</b>
<b>IV ПРЕГЛЕД ДОКТОРСKE ДИСЕРТАЦИЈЕ:</b>
<p>Навести кратак садржај са назнаком броја страница, поглавља, слика, схема, графикона и сл.</p> <p>Докторска дисертација написана је на српском језику и има 179 страна, 9 поглавља, 58 референци, 68 слика, 16 табела и 4 листинга алгоритама. Кључна документација написана је на српском и енглеском језику. Дисертација садржи следећа поглавља:</p> <ol style="list-style-type: none"><li>1. Увод</li><li>2. Машинско учење</li><li>3. Развој алгоритама за кластеровање кернела унутар конволуционих слојева</li><li>4. Побољшање постојећег алгорита за резивање CNN-ова са акцентом на FPGA платформе</li><li>5. Синергија побољшаног алгорита за резивање и кластеровања</li><li>6. Архитектура акцелератора</li><li>7. Функционална верификација и хардверско тестирање акцелератора</li><li>8. Експериментални резултати</li><li>9. Закључак</li><li>10. Референце</li></ol>

## **V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:**

Комисија сматра да је наслов докторске дисертације јасно и прецизно формулисан и да јасно указује на предмет истраживања и садржај рада у оквиру дисертације.

### Поглавље 1

У уводном поглављу јасно је дата мотивација за спроведена истраживања. Такође су таксативно наведени и укратко објашњени главни научни доприноси које је кандидат остварио током рада на теми докторске дисертације.

Комисија сматра да је изложена мотивација за извршена истраживања адекватна и смислена, као и да остварени научни доприноси јасно показују способност кандидата за будући самостални научно-истраживачки рад.

### Поглавље 2

У другом поглављу презентован је кратак увод у област машинског учења. Дато је неколико неформалних дефиниција машинског учења, дефинисани су основни типови машинског учења и описани су основни кораци приликом формирања предиктивног модела помоћу технике надгледаног учења. Затим је дат приказ конволуционих неуронских мрежа, чија је ефикасна акцелерација главна тема докторске дисертације. Након тога дат је исцрпан преглед актуелног стања у области хадверске акцелерације конволуционих неуронских мрежа.

Комисија сматра да је увод у машинско учење и конволуционе неуронске мреже написан на такав начин да је и особама које се раније нису бавиле овим областима могуће разумевање основних принципа. Поред тога, комисија сматра да је приказани преглед актуелног стања у области врло детаљно написан и да обухвата све битне претходно развијене хадверске акцелераторе конволуционих неуронских мрежа.

### Поглавље 3

Треће поглавље презентује први оригинални научни допринос докторске дисертације, поступак кластеровања кернела унутар конволуционих неуронских мрежа. На почетку овог поглавља изложена је мотивација за кластеровање кернела, редукција потребних хадверских ресурса за имплементирање процесирајућих елемената са могућношћу прескакања небитних операција унутар хадверског акцелератора. Затим је детаљно приказан развијени алгоритам за кластеровање кернела, као и алгоритам за преслагање кластерованих кернела.

Комисија је мишљења да је приказан алгоритам описан на нивоу који омогућава његово разумевање, као и евентуалну реимплементацију.

### Поглавље 4

У четвртном поглављу дата је врло јасна аргументација за развој новог алгоритма за орезивање кернела унутар конволуционих неуронских мрежа. Предложени алгоритам за орезивање спада у групу “architecture-aware” алгоритама за орезивање, који приликом орезивања воде рачуна о конкретној архитектури хадверског акцелератора који ће бити коришћен за акцелерацију орезаних конволуционих неуронских мрежа. Опште је познато да се узимањем у обзир карактеристика хадверског акцелератора приликом орезивања конволуционих неуронских мрежа могу постићи знатна побољшања у погледу ефикасности акцелерације, односно потребних хадверских ресурса за имплементацију хадверског акцелератора. Међутим, већина предложених алгоритама за “architecture-aware” орезивање концентришу се на ASIC хадверске акцелераторе, и нису у стању да искористе специфичности FPGA имплементације. Ова чињеница је примећена од стране кандидата, који је извршио њихово прилагођење тако да дају значајно боље резултате када се планира FPGA имплементација хадверског акцелератора конволуционих неуронских мрежа. Овај алгоритам уједно представља други оригинални научни допринос докторске дисертације.

### Поглавље 5

У петом поглављу детаљно је изложен развијени алгоритам који комбинује прилагођени алгоритам за “architecture-aware” орезивање, намењен за коришћење у случајевима када се хадверски акцелератор конволуционих неуронских мрежа имплементира помоћу FPGA технологије, заједно са претходно приказаним алгоритмом за кластеровање кернела. Ово уједно представља трећи оригинални научни допринос докторске дисертације. У овом поглављу такође су приказани и

резултати експеримената који су имали за циљ да испитају перформансе орезивања предложеног алгоритма на неколико стандардних архитектура конволуционих неуронских мрежа. На крају, детаљно су дискутовани утицај предложеног алгоритма за орезивање на хардверске ресурсе потребне за имплементацију акцелератора конволуционих неуронских мрежа, као и на његове перформансе.

Комисија сматра да је приказ алгоритма за кластеровање и “architecture-aware” орезивање кернела врло детаљно изведен, као и да приказани резултати експерименталне анализе његових перформанси јасно показују његову ефикасност.

#### Поглавље 6

Шесто поглавље посвећено је приказу архитектуре хардверског акцелератора конволуционих неуронских мрежа која је пројектована тако да у потпуности искористи предности које се појављују када се конволуционе неуронске мреже орезају са новоразвијеним алгоритмом за орезивање са кластеровањем, представљеним у петом поглављу. Ова архитектура уједно представља четврти оригинални научни допринос докторске дисертације. На почетку поглавља детаљно су образложене оптимизације које су спроведене са циљем што ефикасније акцелерације различитих типова конволуционих слојева унутар конволуционе неуронске мреже. Затим је изложен врло детаљан приказ развијене архитектуре, прво на највишем нивоу хијерархије, а затим и на нивоу свих главних модула. На крају поглавља дат је приказ начина на који је могуће извршити ефикасно скалирање развијене архитектуре, у циљу повећања остваривих перформанси. Комисија констатује да је приказ архитектуре развијеног акцелератора извршен са изузетним нивоом детаља, уз обилато коришћење одличних блок дијаграма и приказа таласних облика карактеристичних сигнала. Све ово омогућује јасан и потпуни увид у детаље функционисања предложене архитектуре.

#### Поглавље 7

Седмо поглавље бави се проблематиком функционалне верификације развијеног хардверског акцелератора. Након кратког теоријског увода у област и значај функционалне верификације приказан је опис развијеног верификационог окружења које је коришћено за верификацију хардверског акцелератора конволуционих неуронских мрежа, дискутованог у шестом поглављу. Интересантно је да је развијено верификационо окружење заправо написано у С програмском језику, што представља одступање од устаљене праксе коришћења специјализованих језика за верификацију хардвера, као што су SystemVerilog или е језик. Међутим разлози због којих се аутор одлучио за овакав избор су више него убедљиви, јер је на овај начин могуће поновно коришћење написаног С кода приликом тестирања и самог коришћења акцелератора на FPGA колу. У наставку поглавља дат је врло детаљан приказ самог тока верификације, илустрован низом конкретних примера. На самом крају поглавља описан је и поступак који је коришћен за тестирање коректног рада развијеног акцелератора на самој FPGA платформи.

Комисија сматра да је кандидат посветио дужну пажњу проблему функционалне верификације једног сложеног дигиталног система, као што је развијени хардверски акцелератор конволуционих мрежа. На основу изложених примера карактеристичних сценарија који су коришћени приликом верификације, који су праћени приказом и детаљним објашњењем таласних облика добијених помоћу симулатора, комисија је сагласна у констатацији да спроведена верификација превазилази стандарде једног истраживачког рада и ближа је стандардима који се срећу у индустријским пројектима.

#### Поглавље 8

Осмо поглавље посвећено је приказу резултата који су добијени спровођењем низа различитих експеримената чији циљ је био да се естимирају захтеви за хардверским ресурсима и перформансе развијеног акцелератора конволуционих неуронских мрежа, као и да се оне упореде са постојећим решењима. На почетку приказани су резултати хардверске имплементације развијеног акцелератора помоћу FPGA кола. Извршена је анализа скалабилности предложене архитектуре, а затим су потребни хардверски ресурси за имплементацију предложеног акцелератора упоређени са потребним хардверским ресурсима за имплементацију већег броја претходно предложених акцелератора конволуционих неуронских мрежа. Затим је дат приказ перформанси три различите конфигурације предложеног акцелератора, а након тога су остварене перформансе упоређене са

перформансама претходно предложених акцелератора, по читавом низу различитих критеријума. Комисија је мишљења да су спроведени експерименти добро осмишљени и постављени, и као такви сасвим адекватни за објективну процену остварених карактеристика развијеног акцелератора. Сами резултати извршених експериментата, по мишљењу чланова комисије, јасно потврђују супериорност развијеног акцелератора у односу на раније предложена решења, која су коришћена за поређење.

#### Поглавље 9

Девето поглавље садржи закључак, у којем су сажето приказани најважнији резултати до којих је аутор дошао приликом израде докторске дисертације.

#### Поглавље 10

Десето поглавље садржи листу референци које су коришћене у докторској дисертацији. Комисија сматра да је приложена листа врло исцрпна и да одговара тематици докторске дисертације.

На основу анализе и вредновања појединих делова докторске дисертације, комисија позитивно оцењује све делове докторске дисертације.

### **VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ:**

Таксативно навести називе радова, где и када су објављени. Прво навести најмање један рад објављен или прихваћен за објављивање у складу са *Правилима докторских студија Универзитета у Новом Саду* који је повезан са садржајем докторске дисертације. У случају радова прихваћених за објављивање, таксативно навести називе радова, где и када ће бити објављени и приложити потврду уредника часописа о томе.

Рад који представља главне доприносе докторске дисертације:

**M23** - Рад у међународном часопису

1. **D. Rakanović**, V. Vranjković, R. Struharik, "Argus CNN Accelerator Based on Kernel Clustering and Resource-Aware Pruning", *Elektronika in Elektrotehnika*, Volume 27, Number 3, June 2021, pp. 57-70, ISSN: 1392-1215, DOI: 10.5755/j02.eie.28922

Остали радови који су уско повезани са темом докторске дисертације:

**M23** - Рад у међународном часопису

1. R. Struharik, B. Vukobratović, A. Erdeljan, **D. Rakanović**, "CoNNA – hardware accelerator for Compressed Convolutional neural networks", *Microprocessors and Microsystems*, Volume 73, March 2020, pp. 1-28, ISSN: 0141-9331, DOI: 10.1016/j.micpro.2020.102991

**M33** – Саопштење са међународног скупа штампано у целини

1. R. Struharik, B. Vukobratović, A. Erdeljan, **D. Rakanović**, "CoNNA – Compressed CNN Hardware Accelerator", 2018 Euromicro Conference on Digital System Design (DSD), Prague, Czech Republic, August 29 - 31 2018., DOI: 10.1109/DSD.2018.00070
2. **D. Rakanović**, A. Erdeljan, V. Vranjković, B. Vukobratović, P. Teodorović, R. Struharik, "Reducing off-chip memory traffic in Deep CNNs using Stick Buffer Cache", 25th Telecommunications Forum TELFOR 2017, Belgrade, Serbia, November 21-22 2017., DOI: 10.1109/TELFOR.2017.8249398

### **VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА:**

Главни резултат предложене докторске дисертације је истовремени развој новог алгоритма за резивање конволуционих неуронских мрежа и хардверског акцелератора, названог Аргус, намењеног за употребу у ембедед системима базираним на FPGA колима. У докторској дисертацији јасно је показано да се упоредним, а не независним, развојем алгоритма за резивање и саме архитектуре акцелератора може постићи веће искоришћење расположивих хардверских ресурса који су присутни на FPGA чиповима, а самим тим и постизање бољих перформанси од претходно

публикованих решења. Развијени алгоритам за орезивање конволуционих неуронских мрежа базира се на увођењу пажљиво одабраних образаца позиција нултих елемената у кернел мапама. Овакав приступ резултује у значајном смањењу хардверских ресурса потребних за имплементацију логике за прескакање непотребних операција приликом процесирања орезаних конволуционих неуронских мрежа. У случају имплементације акцелератора помоћу FPGA технологије показано је да се постиже смањење неопходних хардверских ресурса од два пута.

Додатано, увођењем идеје кластеровања кернела, код које се нулти коефицијенти налазе на истим позицијама унутар свих кернела који чине један кластер, могуће је додатно смањити хардверске ресурсе потребне за имплементацију акцелератора, без значајног смањења тачности у односу на оригиналну конволуциону неуронску мрежу.

Пројектовани Аргус акцелератор користи све наведене бенефите предложеног алгоритма за орезивање конволуционих неуронских мрежа са кластеровањем, те због тога балансирано искоришћава расположиве хардверске ресурсе, а све у циљу побољшања перформанси на компактним FPGA чиповима.

У оквиру докторске дисертације приказани су резултати имплементације Аргус хардверског акцелератора за три различите конфигурације са прогресивно већом снагом процесирања конволуционих неуронских мрежа. Показано је да се перформансе акцелератора врло добро скалирају са повећањем коришћених хардверских ресурса.

У дисертацији су такође приказани и резултати експеримената који су имали за циљ да омогуће поређење перформанси Аргус акцелератора са неким од раније предложених решења. Показано је да Аргус акцелератор на равномеран начин троши различите хардверске ресурсе који су присутни у савременим FPGA колима, те да га је стога могуће ефикасно имплементирати како на најмањим тако и на највећим FPGA чиповима, што није случај са великом већином постојећих решења. Упоредна анализа апсолутних перформанси, исказаних у терминима времена класификације индивидулане улазне инстанце, јасно је показала да су резултати које постиже Аргус акцелератор упоредиви или бољи од перформанси достижних коришћењем претходно публикованих решења.

Додатно, перформансе Аргус акцелератора упоређиване су са перформансама претходно публикованих решења и коришћењем других метрика, исказаних у терминима достигнуг броја операција у јединици времена (исказаних у GOP/s), рачунатих по јединици расположивих и искоришћених ресурса присутних у FPGA колима (LUT, DSP, BlockRAM). На основу презентованих резултата може се приметити да се Аргус акцелератор налази на првом или другом месту без обзира на коришћену метрику.

### **VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА:**

Експлицитно навести позитивну или негативну оцену начина приказа и тумачења резултата истраживања.

Прегледом докторске дисертације Комисија закључује да је приказ дисертације јасно структуриран, и да је тумачење добијених резултата јасно, истраживачки коректно и разумљиво. Сви добијени резултати стављени су у добар контекст и упоређени са до сада познатим и публикованим резултатима. Резултати су приказани исцрпно и прегледно, уз ослањање и навођење претходних истраживачких резултата у овој области.

Дисертација је проверена у софтверу за детекцију плагијаризма (iThenticate). Извештај о подударности је показао је да је преклапање текста у дозвољеним границама.

На основу свега наведеног, комисија **позитивно оцењује** начин приказа и тумачења резултата добијених у оквирима активности кандидата на изради докторске дисертације.

### **IX КОНАЧНА ОЦЕНА ДОКТОРСKE ДИСЕРТАЦИЈЕ:**

Експлицитно навести да ли дисертација јесте или није написана у складу са наведеним образложењем, као и да ли она садржи или не садржи све битне елементе. Дати јасне, прецизне и концизне одговоре на 3. и 4. питање:

1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме?

Дисертација је написана у складу са образложењем наведеним у пријави теме.

2. Да ли дисертација садржи све битне елементе?

Дисертација садржи све битне елементе.

3. По чему је дисертација оригиналан допринос науци?

Оригинални научни допринос докторске дисертације чине четири јасно представљена резултата:

1. Нови алгоритам за кластеровање кернела унутар конволуционих и потпуно повезаних слојева конволуционе неуронске мреже
2. Побољшање постојећег алгоритма за „architecture aware” резивање конволуционе неуронске мреже са акцентом на примену у FPGA колима
3. Нови алгоритам за резивање са кластеровањем који представља синергију алгоритама развијених у тачкама 1. и 2.
4. Развој нове хардверске архитектуре (назване Аргус) која је пројектована тако да искористи све бенефите који се добијају резивањем конволуционих неуронских мрежа коришћењем алгоритма за резивање са кластеровањем развијеним у тачки 3.

Главни резултати добијени током рада на докторској дисертацији објављени су у међународном часопису са импакт фактором, што додатно потврђује оригинални научни допринос докторске дисертације.

4. Који су недостаци дисертације и какав је њихов утицај на резултат истраживања?

Дисертација нема недостатке који би значајније утицали на резултате истраживања.

<b>X ПРЕДЛОГ:</b>
На основу наведеног, комисија предлаже:
<b>а) да се докторска дисертација прихвати, а кандидату одобри одбрана;</b>
б) да се докторска дисертација врати кандидату на дораду (да се допуни односно измени);
в) да се докторска дисертација одбије.

Нови Сад, 16.05.2022. године.

1. Др Иван Мезеи, ванредни професор  
\_\_\_\_\_, председник

2. Др Дејан Вукобратовић, редовни професор  
\_\_\_\_\_, члан

3. Др Тајјана Николић, редовни професор  
\_\_\_\_\_, члан

4. Др Вук Врањковић, ванредни професор  
\_\_\_\_\_, члан

5. Др Растислав Струхарик, редовни професор  
\_\_\_\_\_, ментор

**НАПОМЕНА:** Члан комисије који не жели да потпише извештај јер се не слаже са мишљењем већине чланова комисије, дужан је да унесе у извештај образложење односно разлоге због којих не жели да потпише извештај и да исти потпише.