Multi-Task Deep Learning in the Legal Domain

Christoph Gebendorfer, Garching, 06.08.2018

Software Engineering betrieblicher Informationssysteme (sebis)
Fakultät für Informatik
Technische Universität München

wwwmatthes.in.tum.de

in cooperation with .msg

Agenda

1 Motivation

2 Multi-Task Deep Learning

3 Research Questions

4 Approach

5 Contribution

6 Experiments & Conclusions

Motivation

- Legislative texts
- Regulations
- Enactments
- Patents
- Contracts
- IP documents
- Agreements
- …

Huge amount of unstructured legal documents and text

Demand for **Natural Language Processing**
which needs **annotated** datasets for modelling tasks

Motivation

Annotated legal datasets are highly limited or barely exist at all

- Primarily translation
- Small size
- No testsets

Corpus	Legal	Translation	Classification	Summarization	Size
JRC-Acquis	X	22 languages	X	-	463k docs
DCEP	X	23 languages	-	-	1.5m docs
Europarl	X	20 languages	-	-	30m sen
DGT-TM	X	24 languages	-	-	65m sen
EAC-TM	X	26 languages	-	-	78k sen
MultiUN	X	7 languages	-	-	80m sen
EUbooks	~	26 languages	-	-	173m sen
The HOLJ Corpus	X	english	-	X	188 docs
The Old Bailey	X	english	X	-	1219 docs
ParaCrawl	~	14 languages	-	-	282m sens

Motivation

Text in the legal domain has special properties

- Unique discource type
- Very strict and factual
- References
- Enumerations

What can we do?

Popular methods:

- Creating new datasets
- Use datasets from other domains

What else?

Agenda

Multi-Task Deep Learning

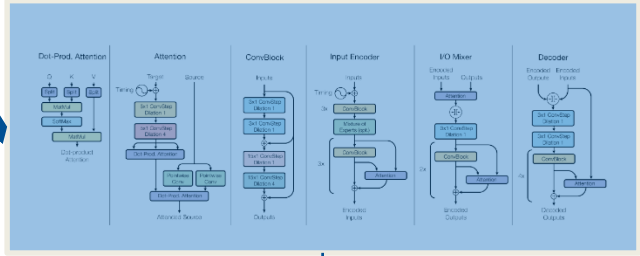
Translation

Summarization

Multi-label Classification

Concurrent Training

MultiModel [1]



Translate Summarize Label

Objective:

- Exploit commonalities and overcome task-specific dataset shortage in the legal domain
- Establish Transfer Learning for better results in legal text tasks
- Support generic / task-independent Deep Learning architecures

Agenda

1. Motivation
2. Multi-Task Deep Learning
3. **Research Questions**
4. Approach
5. Contribution
6. Experiments & Conclusions

Research Questions

1	Can multi-task deep learning be beneficial for tasks in the legal domain?
2	How does training on multiple tasks of the legal domain simultaneously compare to training on each task separately?
3	How far is multi-task deep learning from state-of-the-art solutions in the legal domain?
4	What needs to be considered for choosing suitable hyperparameters for multi-task deep learning in the legal domain?

Agenda

1. Motivation
2. Multi-Task Deep Learning
3. Research Questions
4. **Approach**
5. Contribution
6. Experiments & Conclusions

Approach

Deductive Reasoning

- Search for datasets in the legal domain and process them
- Choose a suitable Multi-Task model
- Integrate datasets into the Multi-Task model
- Conduct experiments
  - Train selected models on special hardware
  - Decode from the trained models
- Evaluate generated information

Backed by literature research

Verify or disprove research questions

Agenda

Contribution

1

6 Ready-to-use Legal Corpora

Multilingual (CS, DE, EN, ES, FR, IT, SV)
- 3 Legal Translation Corpora
- 1 Legal Text Summarization Corpus
- 1 Legal Document Labeling Corpus

German
- 2 Legal Document Classification Corpora

2

Integration into Tensor2Tensor

Problem definitions with data generators
- 35 legal tasks
  - 21 translation (combined translation corpora)
  - 7 summarization
  - 7 multi-label classification
  - 2 classification

Contribution 1

Corpus	Legal	Translation	Classification	Summarization	Size
JRC-Acquis*	X	22 languages	X	X	463k docs
DCEP*	X	23 languages	-	-	1.5m docs
Europarl*	X	20 languages	-	-	30m sen
Legal GCD*	X	german	X	-	42k docs
DGT-TM	X	24 languages	-	-	65m sen
EAC-TM	X	26 languages	-	-	78k sen
MultiUN	X	7 languages	-	-	80m sen
EUbooks	~	26 languages	-	-	173m sen
The HOLJ Corpus	X	english	-	X	188 docs
The Old Bailey	X	english	X	-	1219 docs
ParaCrawl	~	14 languages	-	-	282m sens

--- Processed

*Available online for download at mediaTUM

Contribution 2

Legal Translation Tasks

- CS-DE, CS-EN, CS-ES, CS-FR, CS-IT, CS-SV
- DE-EN, DE-ES, DE-FR, DE-IT, DE-SV
- EN-ES, EN-FR, EN-IT, EN-SV
- ES-FR, ES-IT, ES-SV
- FR-IT, FR-SV
- IT-SV

Legal Summarization Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Legal Multi-Labeling Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Legal Classification Tasks

- Court
- Verdict

Agenda

1. Motivation
2. Multi-Task Deep Learning
3. Research Questions
4. Approach
5. Contribution
6. **Experiments & Conclusions**

Experiments - Setup

		Machine 1	Machine 2	Machine 3 (DGX-1)
GPUs		4x GTX 1080 TI	4x Tesla K80	8x Tesla V100
Cores		~14k	~10k	~41k
Memory		4x 11GB	4x 12GB	8x 16GB
Training Steps	Translation	500k	500k	250k
	Summarization	100k	100k	50k
	Classification	100k	100k	50k
Training Time (dependent on 6.1.2)	Single-Task	25.2 s/100 steps	86.2 s/100 steps	86.4 s/100 steps
	Multi-Task (5 Tasks)	51.8 s/100 steps	-	155.5 s/100 steps

Table 6.1.: Machines used to train the models

	MultiModel Light (MM-L)	MultiModel Base (MM-B)	Transformer Base (TF-B) [2]
Hidden Size	128	512	512
Filter Size	1024	2048	2048
Batch Size	1024	2048	2048
Total Parameters	~61m	~660m	~51m

Table 6.2.: Model hyperparameter sets

Experiments - Setup

msg sebis Azure lrz Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften

		Machine 1	Machine 2	Machine 3 (DGX-1)
GPUs		4x GTX 1080 TI	4x Tesla K80	8x Tesla V100
Cores		~14k	~10k	~41k
Memory		4x 11GB	4x 12GB	8x 16GB
Training Steps	Translation	500k	500k	250k
	Summarization	100k	100k	50k
	Classification	100k	100k	50k
Training Time (dependent on 6.1.2)	Single-Task	25.2 s/100 steps	86.2 s/100 steps	86.4 s/100 steps
	Multi-Task (5 Tasks)	51.8 s/100 steps	-	155.5 s/100 steps

Table 6.1.: Machines used to train the models

Trained on

	MultiModel Light (MM-L)	MultiModel Base (MM-B)	Transformer Base (TF-B)
Hidden Size	128	512	512
Filter Size	1024	2048	2048
Batch Size	1024	2048	2048
Total Parameters	~61m	~660m	~51m

Table 6.2.: Model hyperparameter sets

Experiments - Metrics

Translation

$$BLEU = min(1, \frac{hypothesis\_length}{reference\_length})(\prod_{i=1}^{4} precision_i)^{\frac{1}{4}}$$

$$CHRF_\beta = (1+\beta^2)\frac{chrP \cdot chrR}{\beta^2 \cdot chrP + chrR}, \begin{matrix}\beta = 3 \\ n = 6\end{matrix}$$

Summarization

$$ROUGE_N = \frac{\sum_{S \in reference\_summaries} \sum_{gram_n \in S} count\_match(gram_n)}{\sum_{S \in reference\_summaries} \sum_{gram_n \in S} count(gram_n)}$$

Multi-Labeling

$$Accuracy = \frac{true\_positives}{all\_labels} \qquad Precision = \frac{TP}{TP+FP} \qquad Recall = \frac{TP}{TP+FN}$$

$$Fscore = 2 \bullet \frac{precision \bullet recall}{precision + recall} \qquad At\ least\ 1 = \frac{label\_correct_{\geq 1}}{all\_documents}$$

Experiments - Single-Task Translation

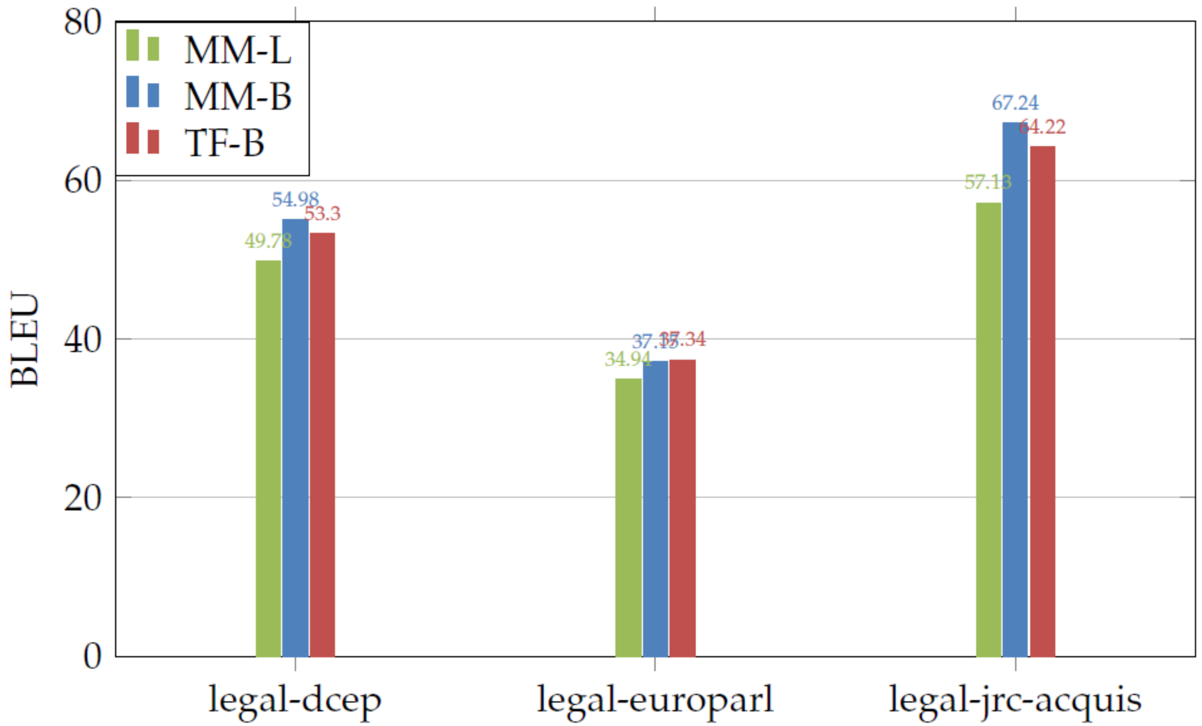


Figure 6.1.: German-to-English single-task translation performance of the MultiModel Light (MM-L), MultiModel Base (MM-B) and Transformer Base (TF-B) - BLEU

Experiments - Single-Task Translation

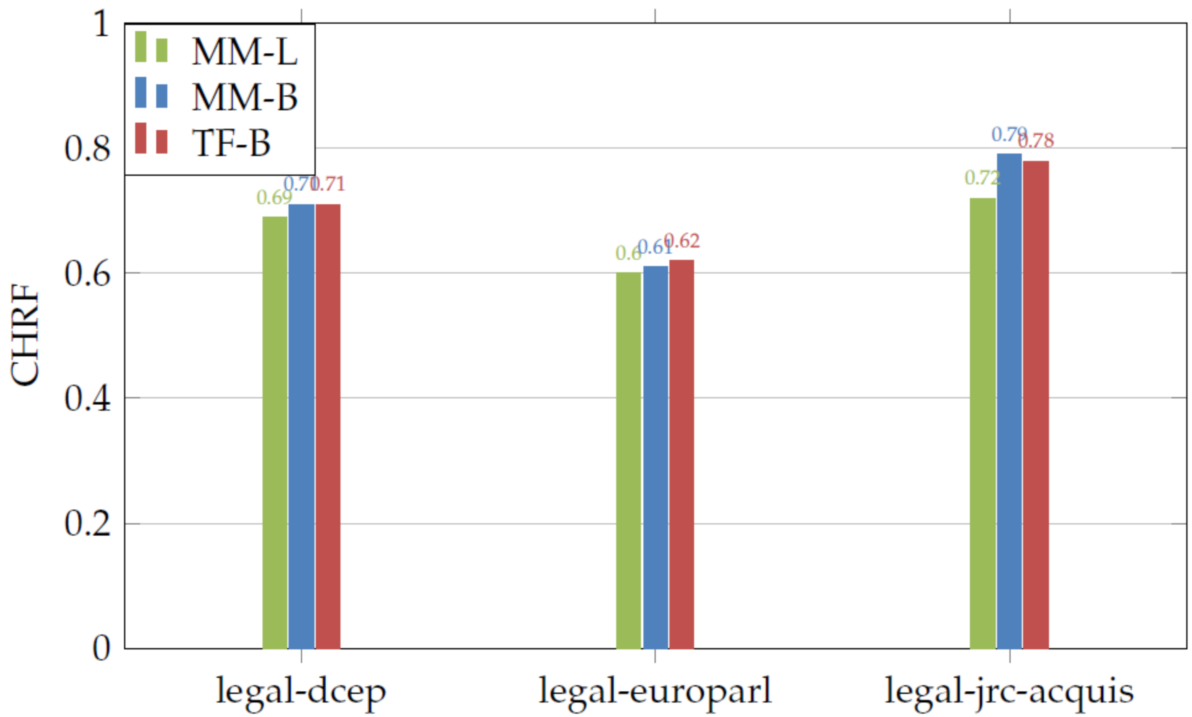


Figure 6.2.: German-to-English single-task translation performance of the MultiModel Light (MM-L), MultiModel Base (MM-B) and Transformer Base (TF-B) - CHRF

Experiments - Single-Task Translation

	BLEU	Example
Input	-	9 . Argentinien gewährleistet die Einhaltung dieser Vereinbarung insbesondere dadurch , daß es innerhalb der in dieser Vereinbarung festgelegten Mengen Ausfuhrlizenzen für die unter Nummer 1 genannten Erzeugnisse erteilt .
MM-L	17.61	9. Argentina shall ensure compliance with this Agreement by granting the export licences referred to in point 1 within the quantities laid down in this Agreement.
MM-B	29.63	9. Argentina shall ensure compliance with this Agreement, in particular by issuing export licences for the products referred to in point 1 within the quantities specified in this Agreement.
TF-B	40.09	9. Argentina shall ensure compliance with this Agreement in particular by issuing export licences for the products referred to in point 1 within the limits of the quantities laid down in this Agreement.
Reference	-	9. Argentina shall ensure that this arrangement is observed, in particular, by issuing export certificates covering the products referred to in paragraph 1 within the limits of the quantities covered by this arrangement.

Table 6.4.: Single-task translation examples of the legal-jrc-acquis by the MultiModel Light (MM-L), MultiModel Base (MM-B) and Transformer Base (TF-B)

Experiments - Single-Task Translation

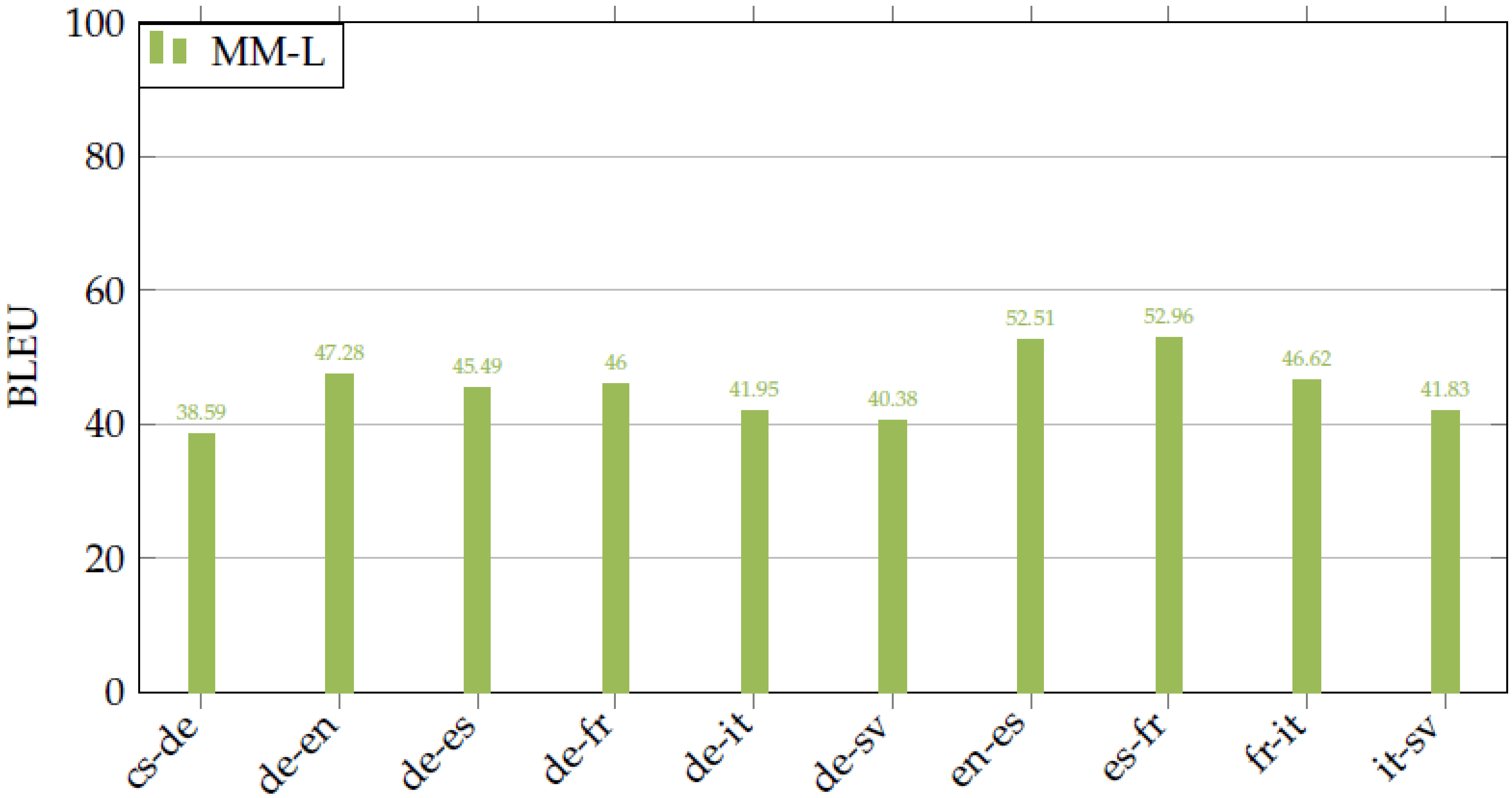


Figure 6.3.: Mean scores across corpora (legal-dcep, legal-europarl, legal-jrc-acquis) of the MultiModel Light (MM-L) - BLEU

Experiments - Single-Task Summarization

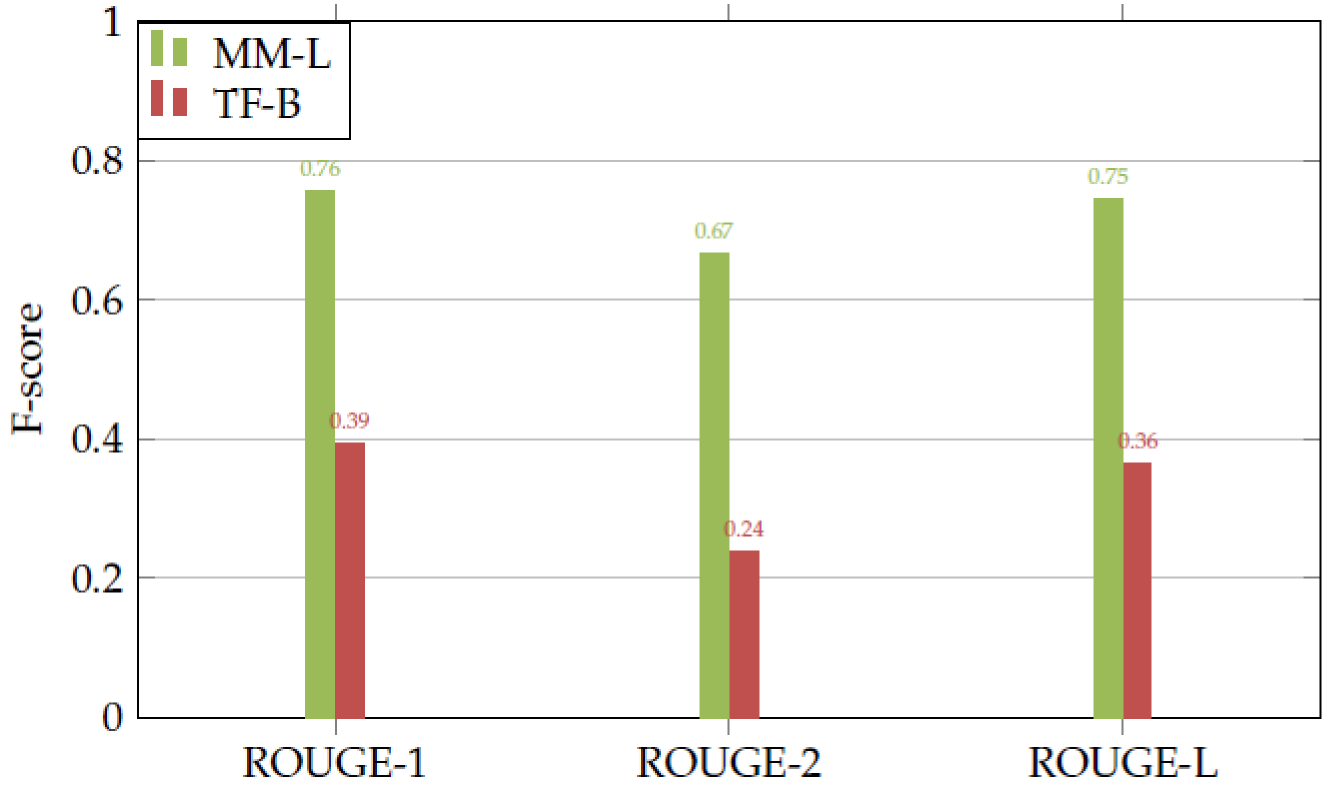


Figure 6.5.: German single-task summarization performance of the MultiModel Light (MM-L) and Transformer Base (TF-B) - F-score

Experiments - Single-Task Multi-Labeling

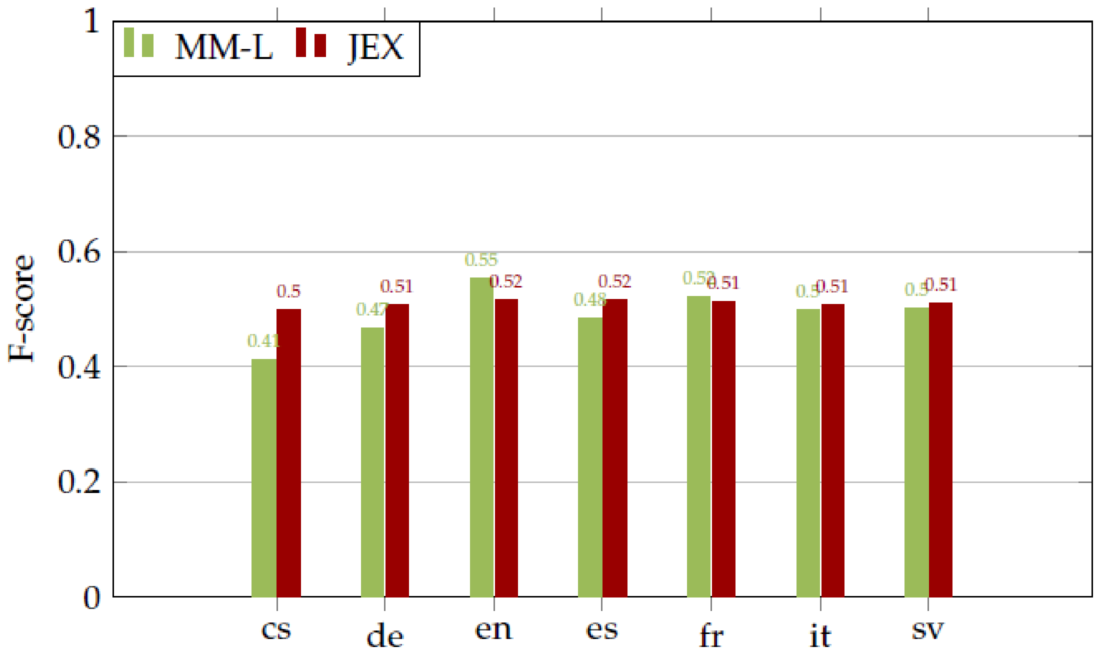


Figure 6.7.: Single-task multi-label classification performance of the MultiModel Light (MM-L) across languages - F-score

		MM-L single	JRC EuroVoc Indexer JEX
cs	Accuracy	0.366	-
	Recall	0.408	**0.521**
	Precision	0.413	**0.469**
	F-score	0.411	**0.493**
	Atleast 1	0.708	-
de	Accuracy	0.422	-
	Recall	0.465	**0.473**
	Precision	0.471	**0.549**
	F-score	0.468	**0.519**
	Atleast 1	0.759	-
en	Accuracy	0.493	-
	Recall	0.543	**0.555**
	Precision	**0.563**	0.480
	F-score	**0.553**	0.523
	Atleast 1	0.854	-
es	Accuracy	0.437	-
	Recall	0.476	**0.555**
	Precision	**0.493**	0.480
	F-score	0.484	**0.519**
	Atleast 1	0.774	-
fr	Accuracy	**0.463**	-
	Recall	0.509	**0.554**
	Precision	**0.532**	0.478
	F-score	**0.520**	0.513
	Atleast 1	**0.845**	-
it	Accuracy	**0.441**	-
	Recall	0.485	**0.546**
	Precision	0.509	0.471
	F-score	0.497	**0.506**
	Atleast 1	0.812	-
sv	Accuracy	**0.438**	-
	Recall	0.483	**0.547**
	Precision	**0.521**	0.479
	F-score	0.501	**0.511**
	Atleast 1	**0.792**	-

Experiments - Single-Task Training

- Baseline
- Base versions of the models outperform light version
- Transformer model performs poorly in summarization and multi-labeling
- Multimodel already reaches state-of-the-art results in single-task training

Experiments - Multi-Task Translation

Joint Translation 5 German Tasks - jt-pool-5

Legal Translation Tasks

- CS-DE, CS-EN, CS-ES, CS-FR, CS-IT, CS-SV
- DE-EN, DE-ES, DE-FR, DE-IT, DE-SV
- EN-ES, EN-FR, EN-IT, EN-SV
- ES-FR, ES-IT, ES-SV
- FR-IT, FR-SV
- IT-SV

Legal Summarization Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Legal Multi-Labeling Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Experiments - Multi-Task Translation

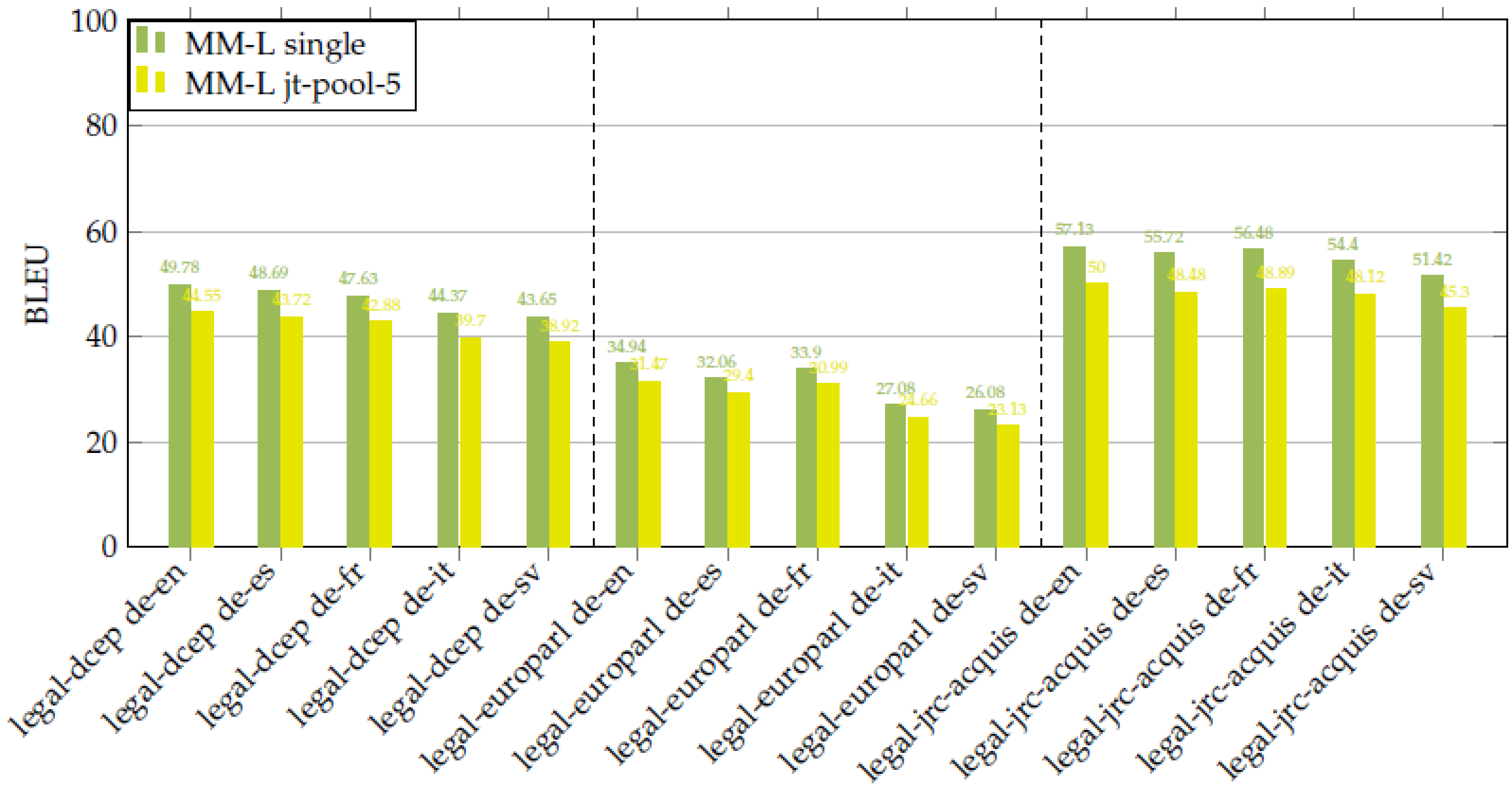


Figure 6.8.: Single-task & multi-task (jt-pool-5) translation performance of the Multi-Model Light (MM-L) - BLEU

Experiments - Multi-Task Translation

Joint Translation 7 Chained Tasks - jt-chain-7

Legal Translation Tasks

- CS-DE, CS-EN, CS-ES, CS-FR, CS-IT, CS-SV
- DE-EN, DE-ES, DE-FR, DE-IT, DE-SV
- EN-ES, EN-FR, EN-IT, EN-SV
- ES-FR, ES-IT, ES-SV
- FR-IT, FR-SV
- IT-SV

Legal Summarization Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Legal Multi-Labeling Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Experiments - Multi-Task Translation

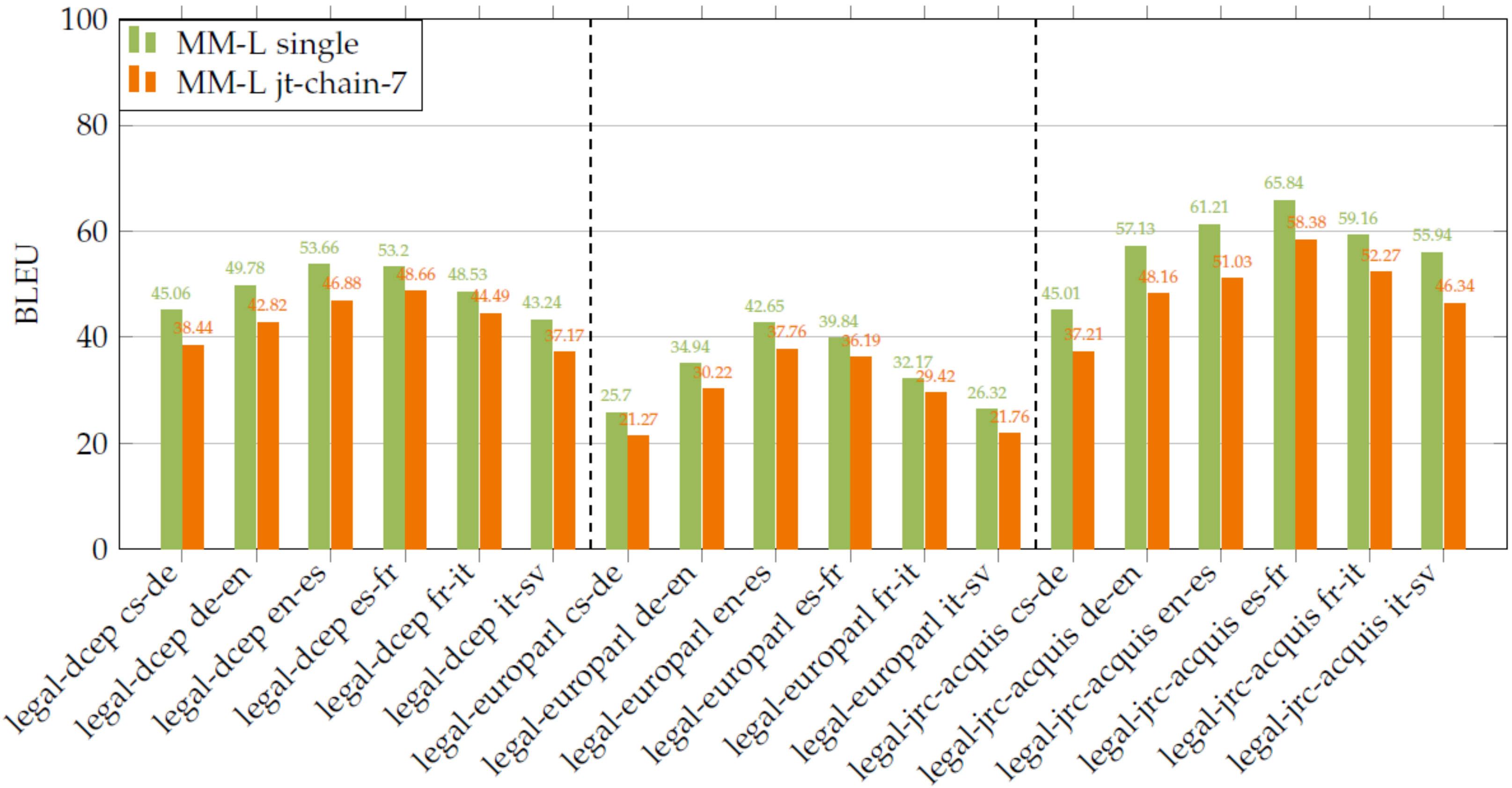


Figure 6.10.: Single-task & multi-task (jt-chain-7) translation performance of the Multi-Model Light (MM-L) - BLEU

Experiments - Multi-Task Translation

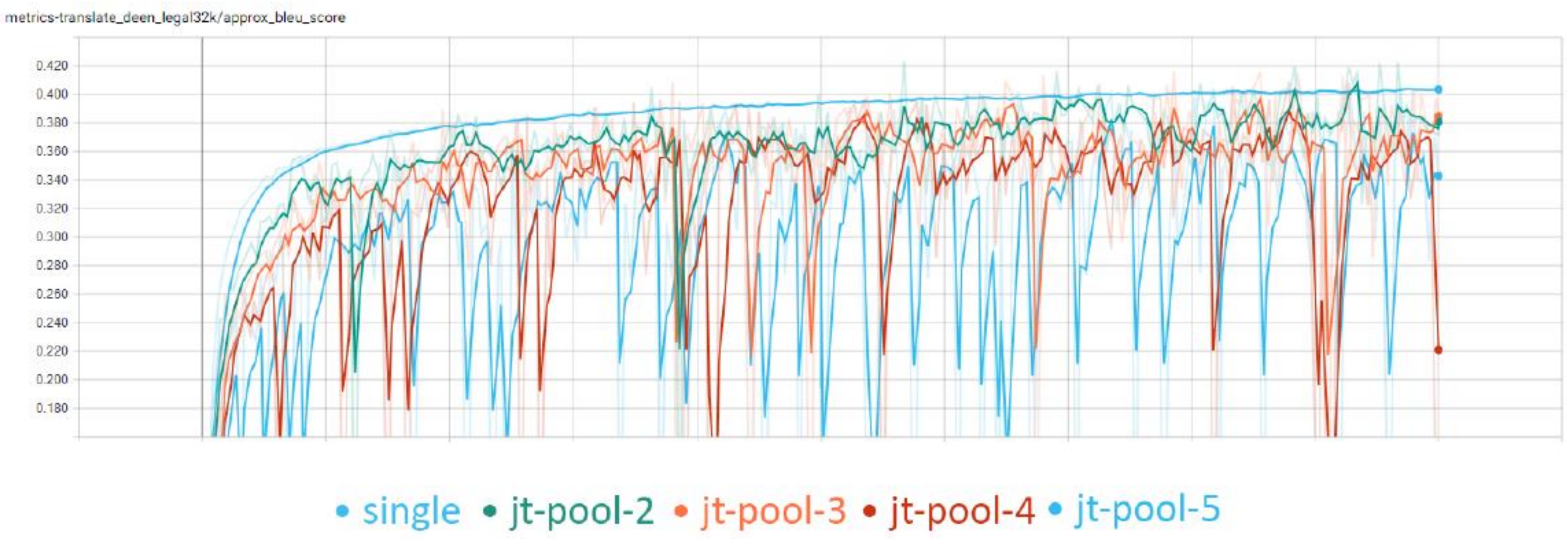


Figure 6.9.: Translation performance depending on the amount of tasks of the MultiModel Light (MM-L) - BLEU

The more tasks are joined together, the worse performs the model

- Transfer Learning does not take place

Insufficient Capacity?

Experiments - Multi-Task Translation

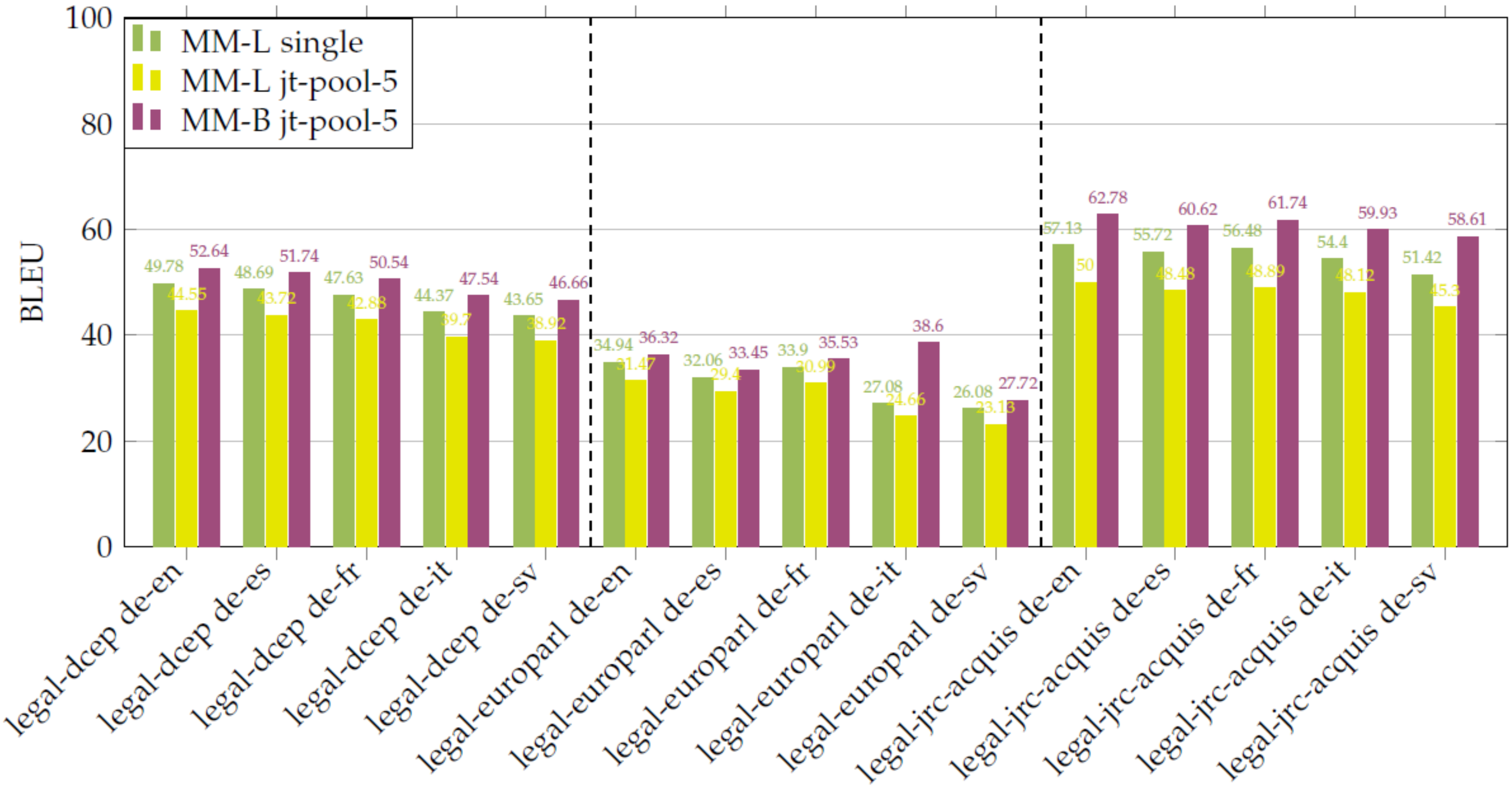


Figure 6.12.: Single-task & multi-task (jt-pool-5) translation performance of the Multi-Model Light - BLEU

- Light version (MM-L) has insufficient capacity

Experiments - Multi-Task Translation

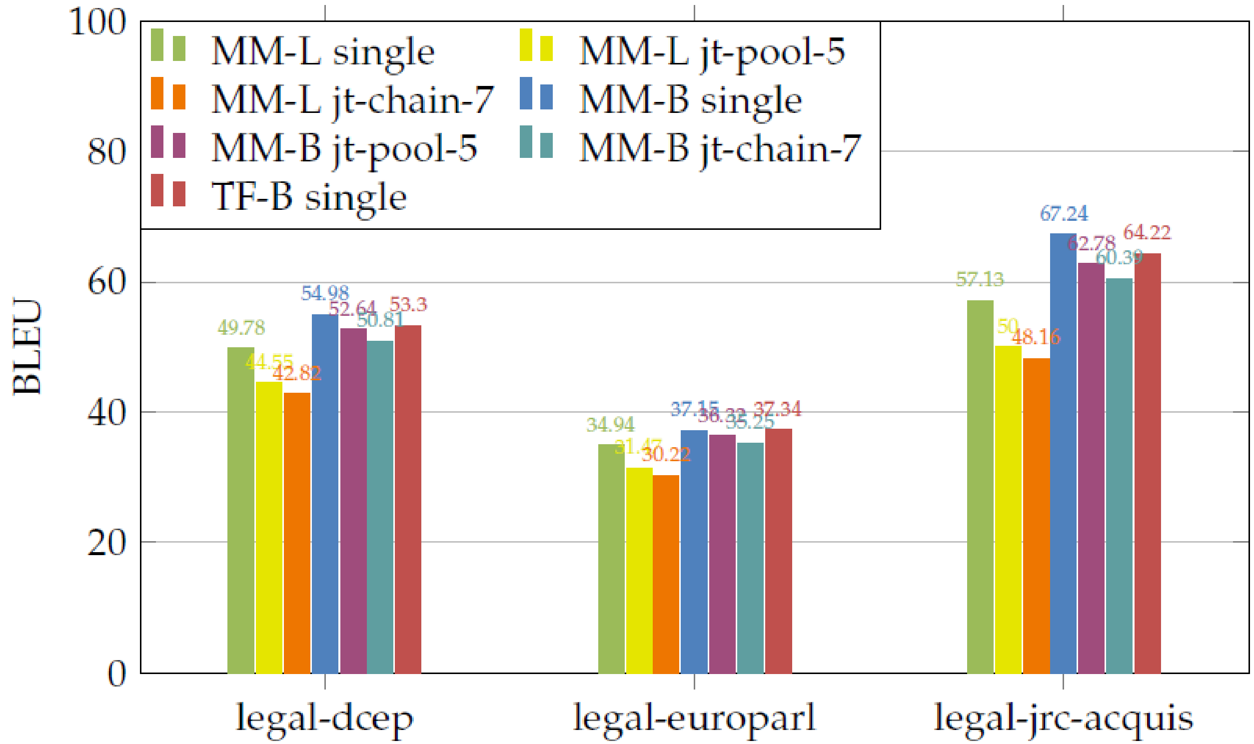


Figure 6.11.: German-to-English translation performance of single-task & multi-task translation combinations trained on the MultiModel Light (MM-L), Multi-Model Base (MM-B) and Transformer Base (TF-B) - BLEU

- Pure translation combinations not that promising

Experiments - Multi-Task Summarization

Joint Summarization 7 Tasks - js-7

Legal Translation Tasks

- CS-DE, CS-EN, CS-ES, CS-FR, CS-IT, CS-SV
- DE-EN, DE-ES, DE-FR, DE-IT, DE-SV
- EN-ES, EN-FR, EN-IT, EN-SV
- ES-FR, ES-IT, ES-SV
- FR-IT, FR-SV
- IT-SV

Legal Summarization Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Legal Multi-Labeling Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Experiments - Multi-Task Summarization

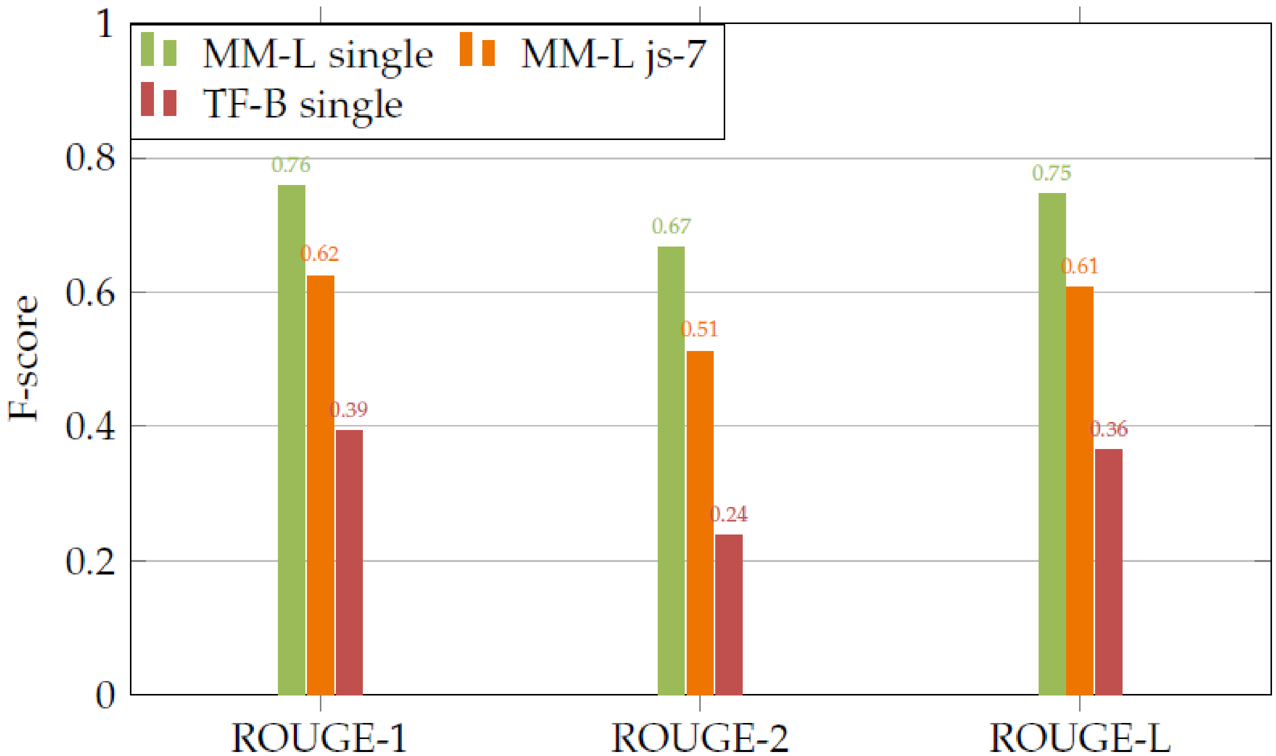


Figure 6.14.: Single-task & multi-task (js-7) summarization performance of the Multi-Model Light (MM-L) and Transformer-Base (TF-B) - BLEU

Experiments - Multi-Task Multi-Labeling

Joint Multi-Labeling 7 Tasks - jl-7

Legal Translation Tasks

- CS-DE, CS-EN, CS-ES, CS-FR, CS-IT, CS-SV
- DE-EN, DE-ES, DE-FR, DE-IT, DE-SV
- EN-ES, EN-FR, EN-IT, EN-SV
- ES-FR, ES-IT, ES-SV
- FR-IT, FR-SV
- IT-SV

Legal Summarization Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Legal Multi-Labeling Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Experiments - Multi-Task Multi-Labeling

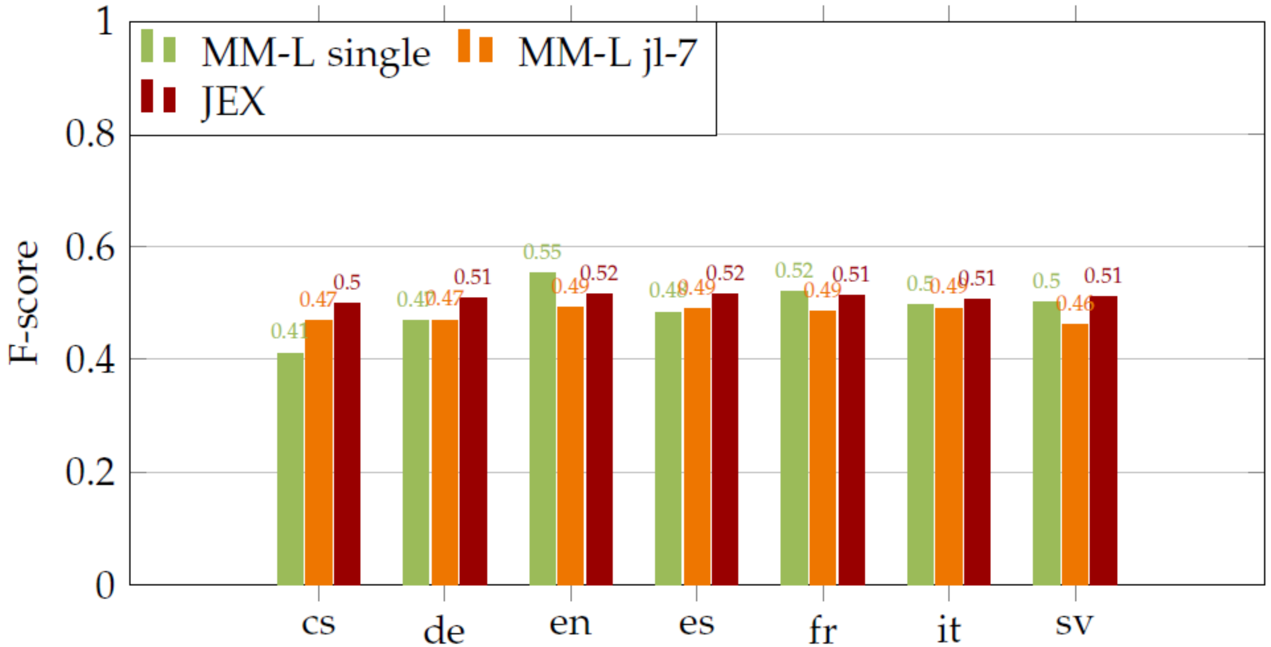
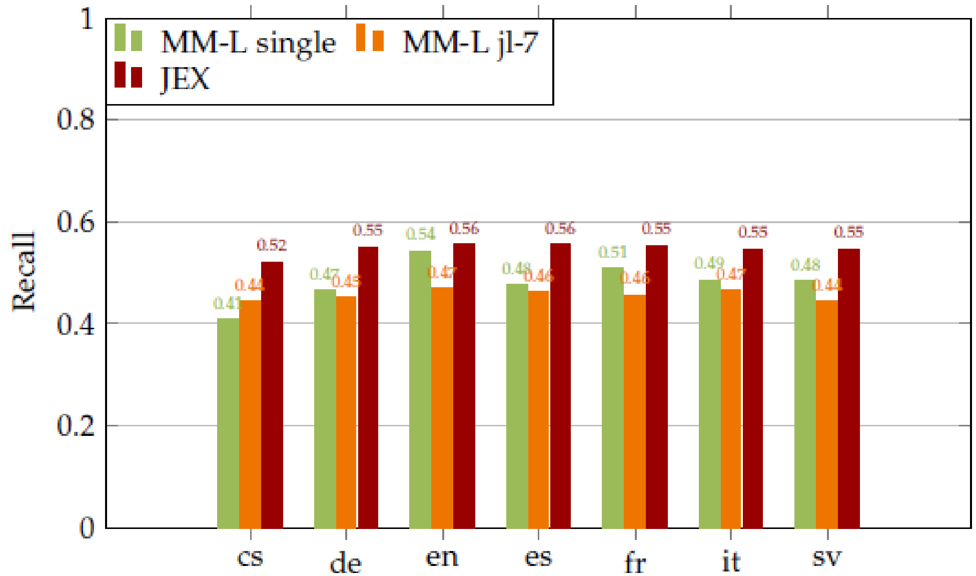
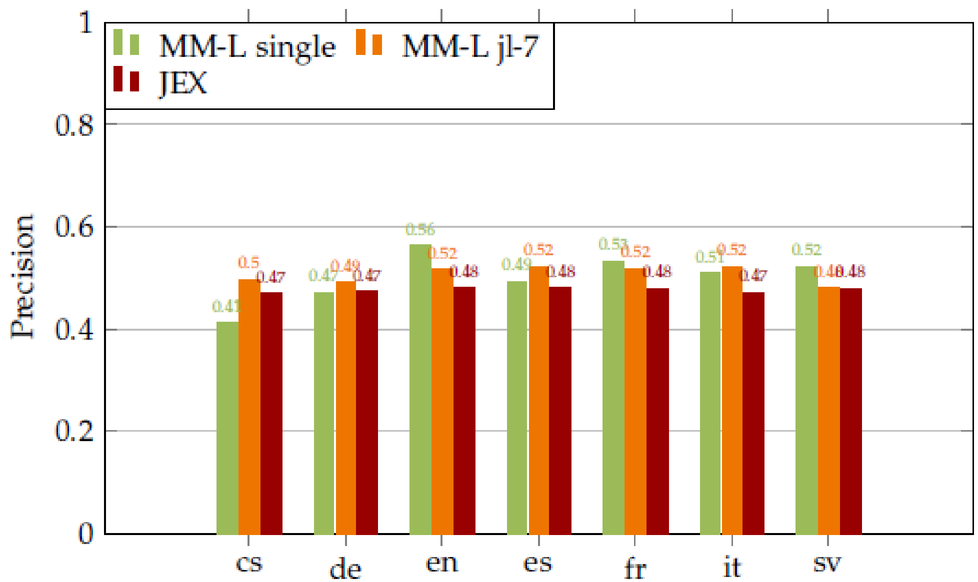


Figure 6.16.: Single-task & multi-task (jl-7) multi-label classification performance of the MultiModel Light (MM-L) and JRC EuroVoc Indexer JEX [3] - F-score

Experiments - Multi-Task Multi-Labeling





Experiments - Multi-Task Across Task Families

Joint Across Task Families 3 Tasks - ja-7

Legal Translation Tasks

- CS-DE, CS-EN, CS-ES, CS-FR, CS-IT, CS-SV
- DE-EN, DE-ES, DE-FR, DE-IT, DE-SV
- EN-ES, EN-FR, EN-IT, EN-SV
- ES-FR, ES-IT, ES-SV
- FR-IT, FR-SV
- IT-SV

Legal Summarization Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Legal Multi-Labeling Tasks

- CS
- DE
- EN
- ES
- FR
- IT
- SV

Experiments - Multi-Task Across Task Families

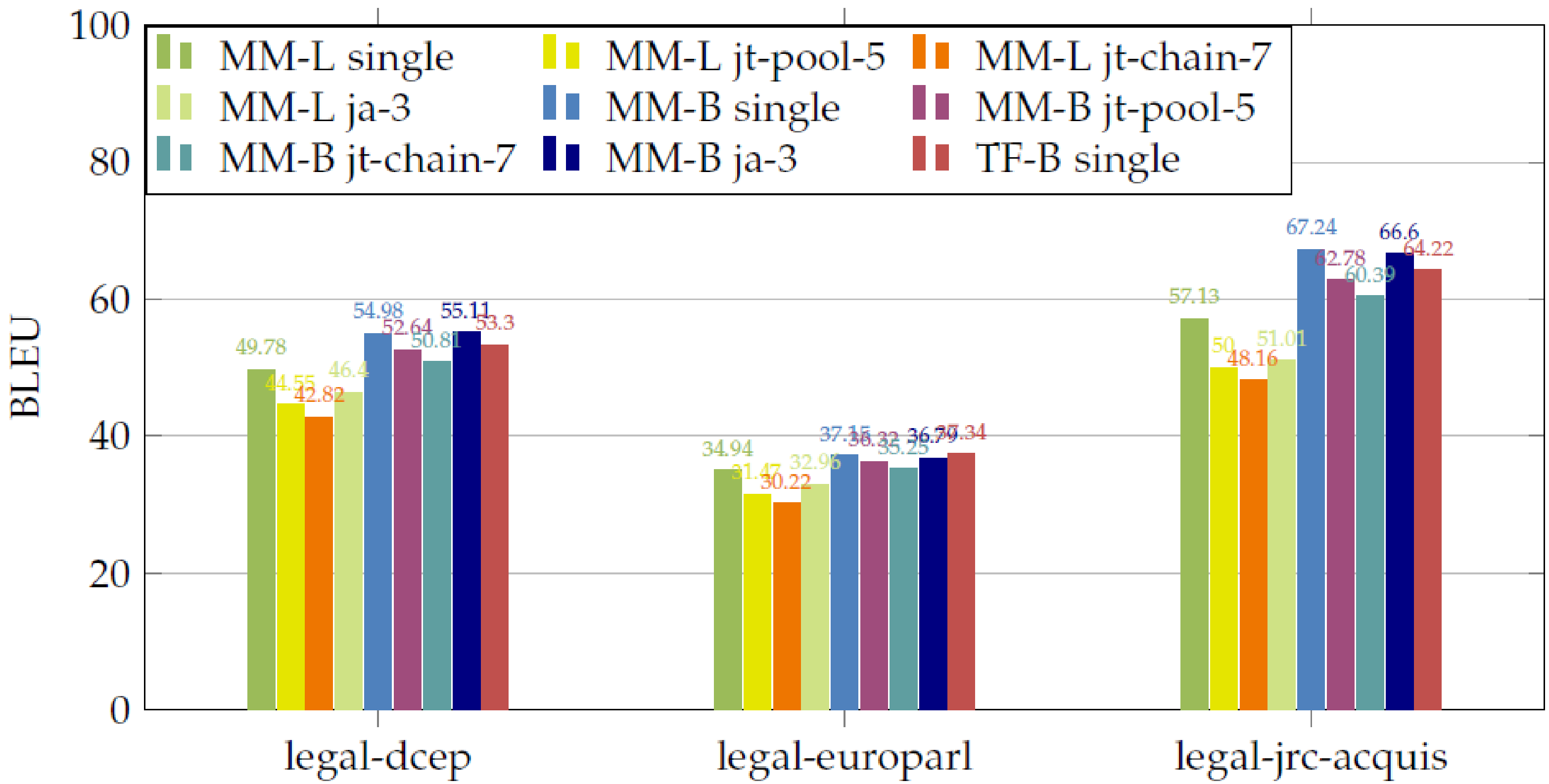


Figure 6.19.: Final German-to-English translation performance of all single-task & multi-task translation combinations trained on the MultiModel Light (MM-L), MultiModel Base (MM-B) and Transformer Base (TF-B) - BLEU

Experiments - Multi-Task Across Task Families

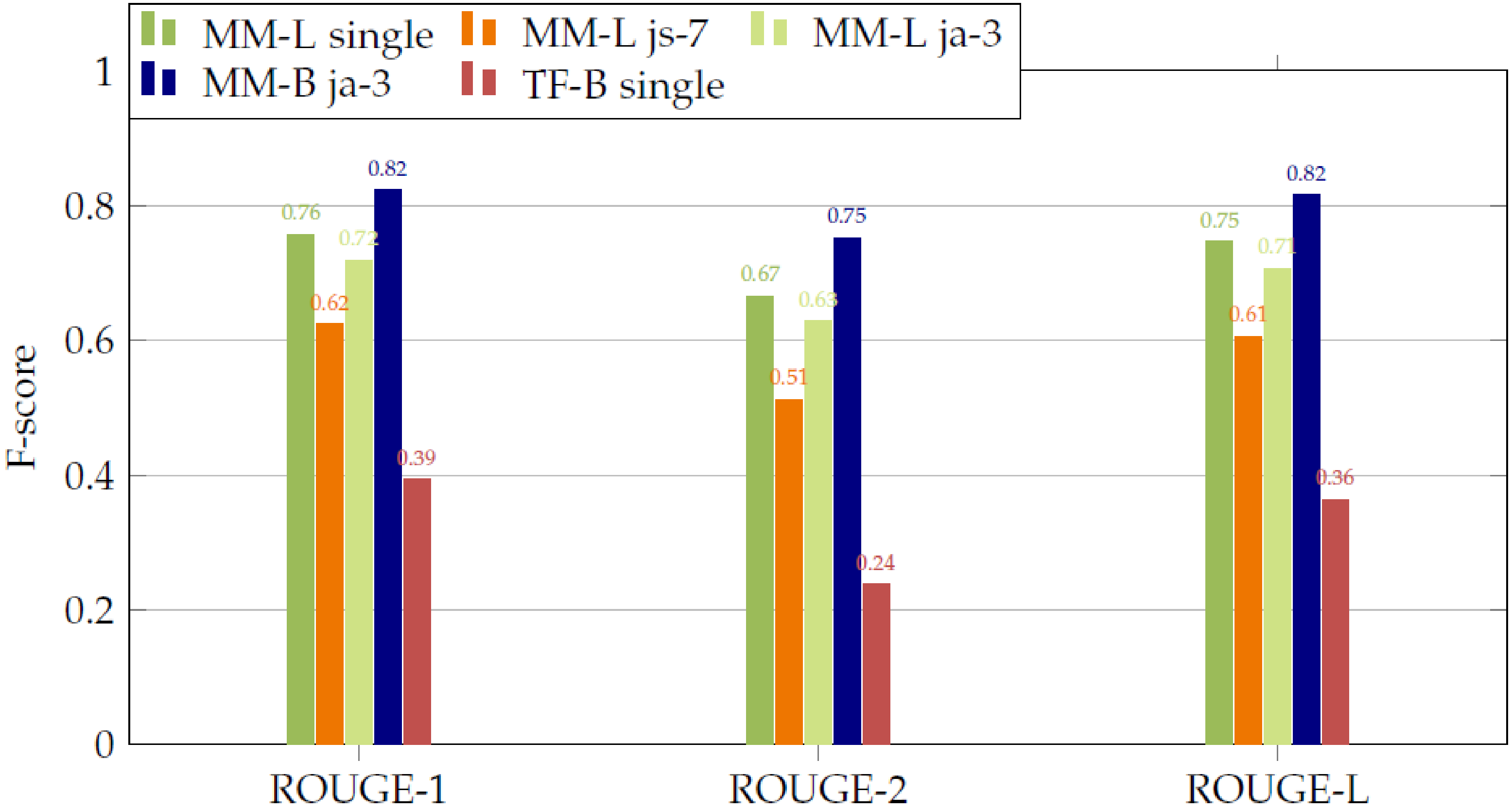


Figure 6.20.: Single-task & multi-task (js-7, ja-3) summarization performance of the MultiModel Light (MM-L), MultiModel Base (MM-B) and Transformer Base (TF-B) - F-score

Experiments - Multi-Task Across Task Families

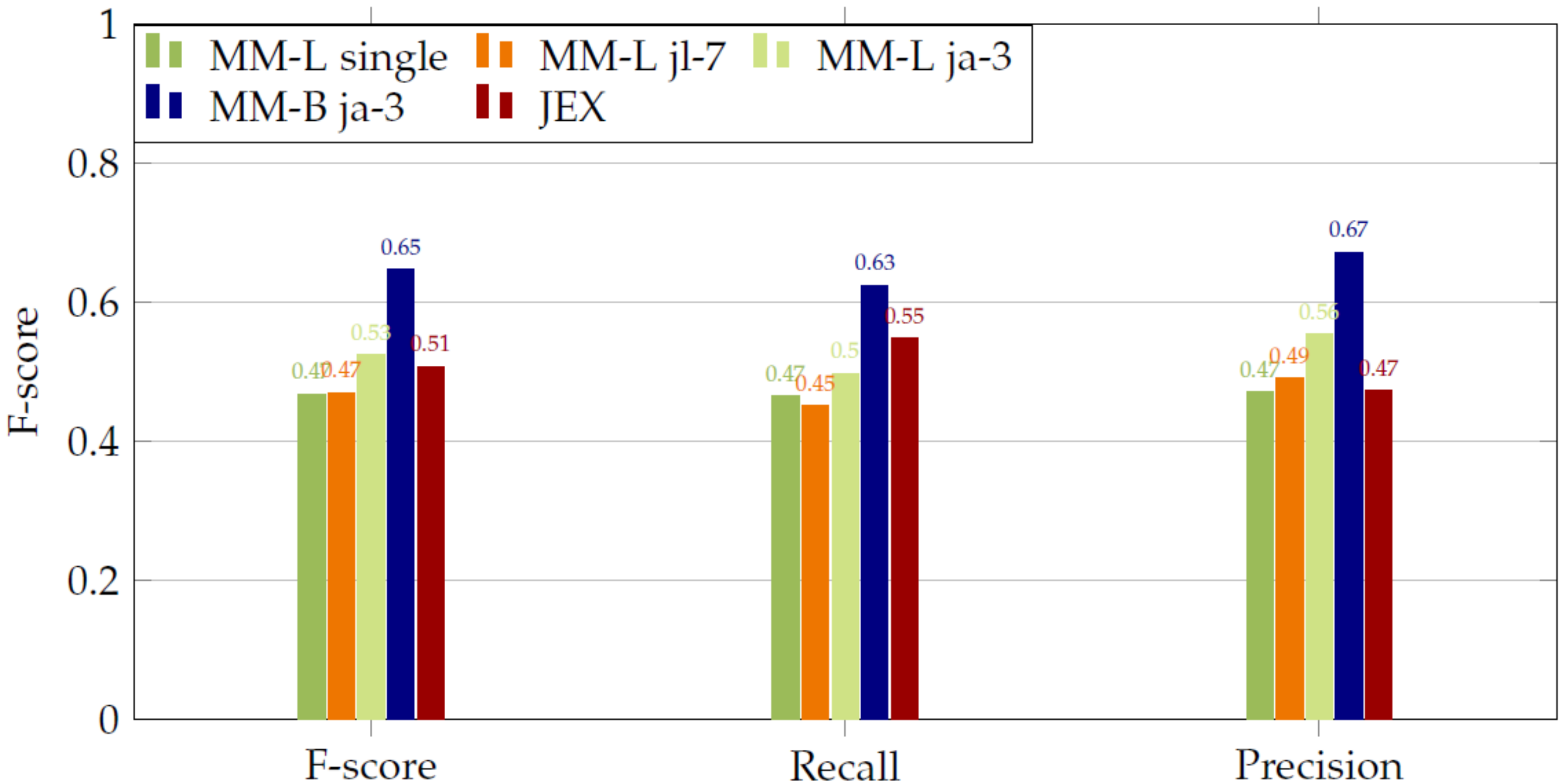


Figure 6.21.: German single-task & multi-task (jl-7, ja-3) multi-label classification performance of the MultiModel Light (MM-L), MultiModel Base (MM-B) and JRC EuroVoc Indexer JEX [3] - F-score, Recall, Precision

Conclusions

1
- …can be beneficial for tasks depending on the task
- …is especially useful where data is sparse

2
- …across task families yields better results than joining inside a task family
- …has shorter training times compared to single-task training

3
- …reaches state-of-the-art results in the legal domain
- …models (at least the MultiModel) are all-rounders

4
- …dataset size must be considered
- …demands high capacity

References

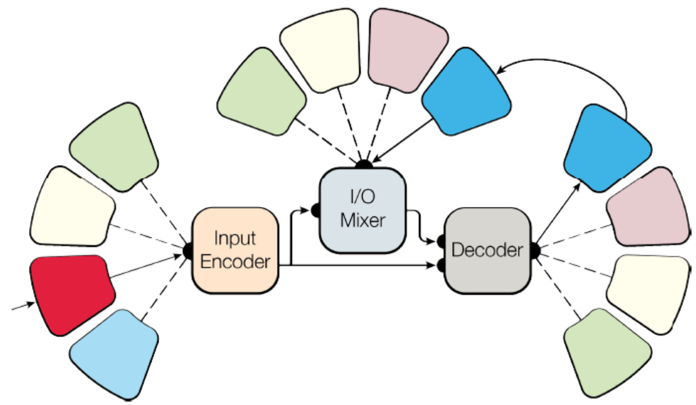
[1] L. Kaiser, A. N. Gomez, N. Shazeer, A. Vaswani, N. Parmar, L. Jones, and J. Uszkoreit, “One model to learn them all,” CoRR, vol. abs/1706.05137, 2017. arXiv: 1706.05137.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” CoRR, vol. abs/1706.03762, 2017. arXiv: 1706.03762.

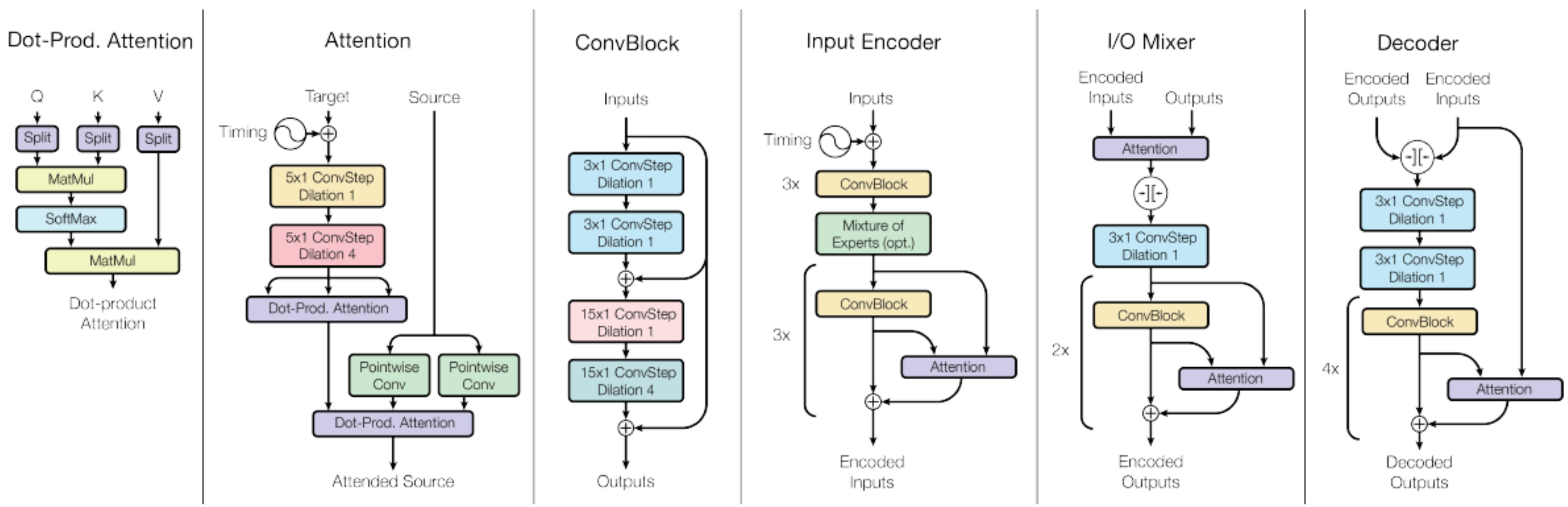
[3] R. Steinberger, M. Ebrahim, and M. Turchi, “JRC eurovoc indexer JEX - A freely available multi-label categorisation tool,” CoRR, vol. abs/1309.5223, 2013. arXiv: 1309.5223.

MultiModel Architecture [1]

Architecture

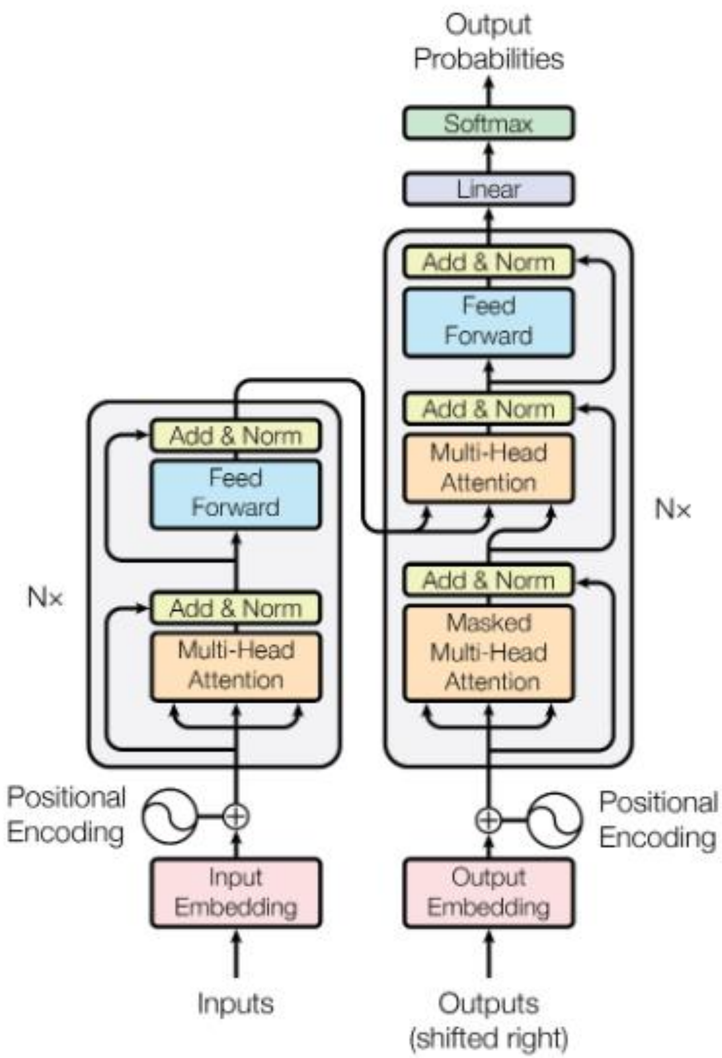


Building Blocks



Transformer Architecture [2]

Architecture



Single-Task vs. Multi-Task

Single-Task

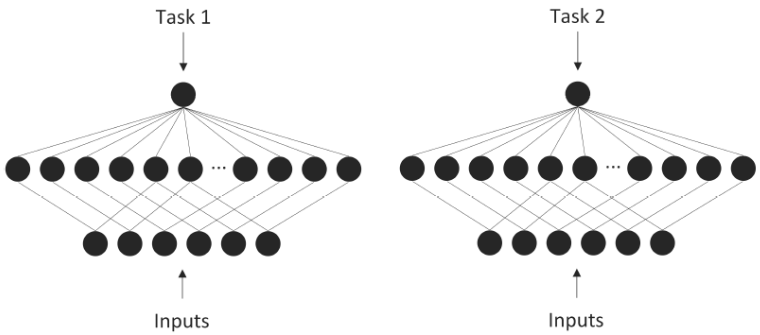


Figure 2.2.: Visualization of single-task learning with two artificial networks on two tasks

Multi-Task

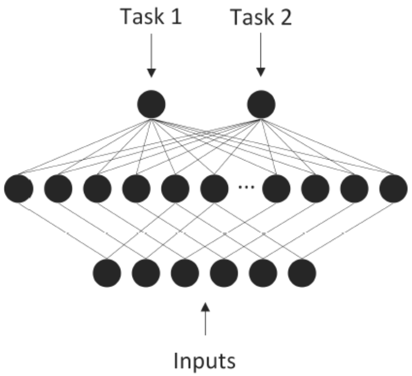
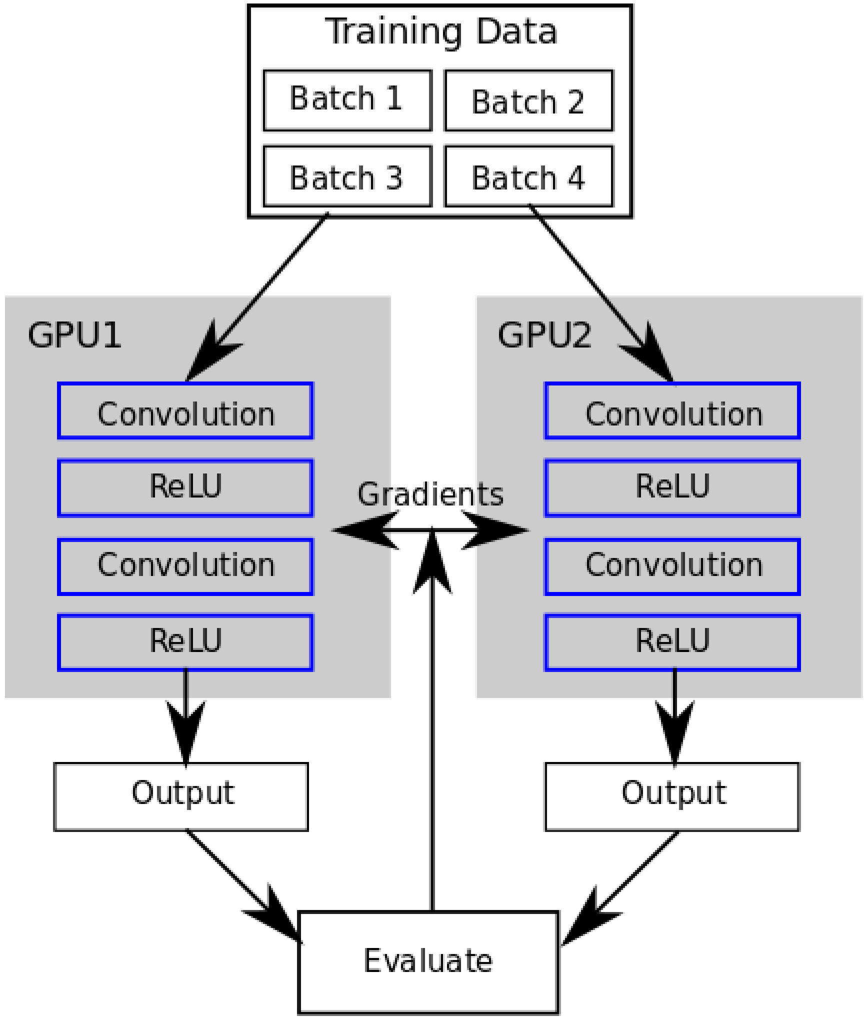


Figure 2.3.: Visualization of multi-task learning with one artificial network on two tasks

Training on multiple GPUs



https://blog.rescale.com/wp-content/uploads/2016/07/dataparallel.png

Download Corpora

Corpus	Type	Link
legal-dcep	Translation	https://mediatum.ub.tum.de/1446648
legal-europarl	Translation	https://mediatum.ub.tum.de/1446650
legal-jrc-acquis	Translation	https://mediatum.ub.tum.de/1446655
legal-jrc-acquis-summarize	Summarization	https://mediatum.ub.tum.de/1446654
legal-jrc-acquis-label	Classification	https://mediatum.ub.tum.de/1446653
legal-gcd, legal-gcd-court & legal-gcd-verdict	Classification	https://mediatum.ub.tum.de/1446651

Table 4.15.: Links to MediaTUM for the download of the legal corpora