



Algorithmen zur Analyse historischer Landkarten

Benedikt Budig
Universität Würzburg

Einführung

Einführung

Algorithmen zur Analyse historischer Landkarten

Einführung

Algorithmen zur Analyse historischer Landkarten



Einführung

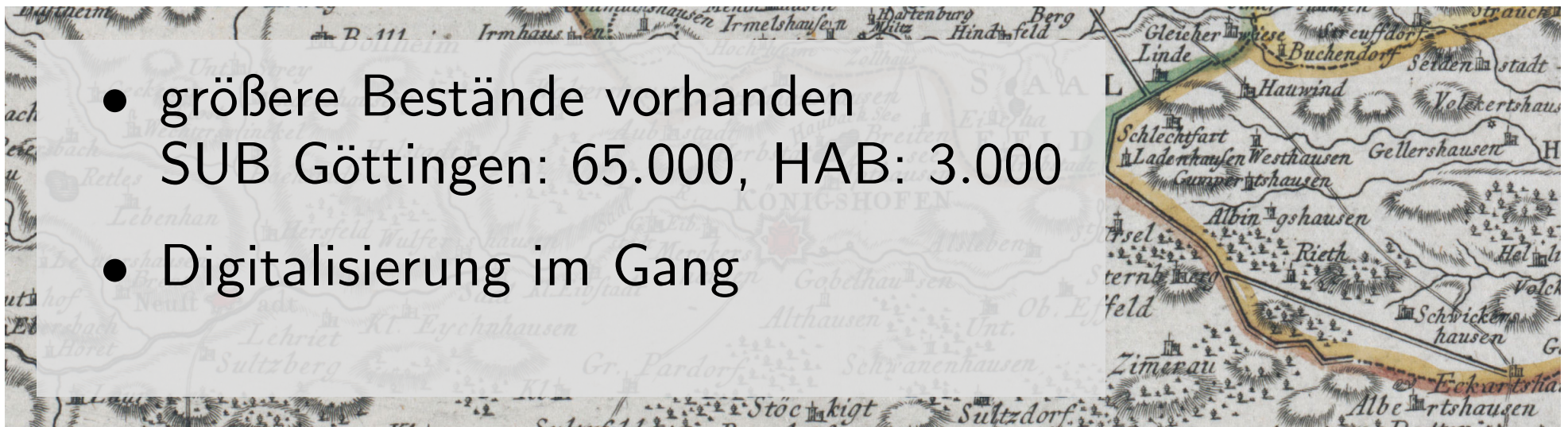
Algorithmen zur Analyse historischer Landkarten



Einführung

Algorithmen zur Analyse historischer Landkarten

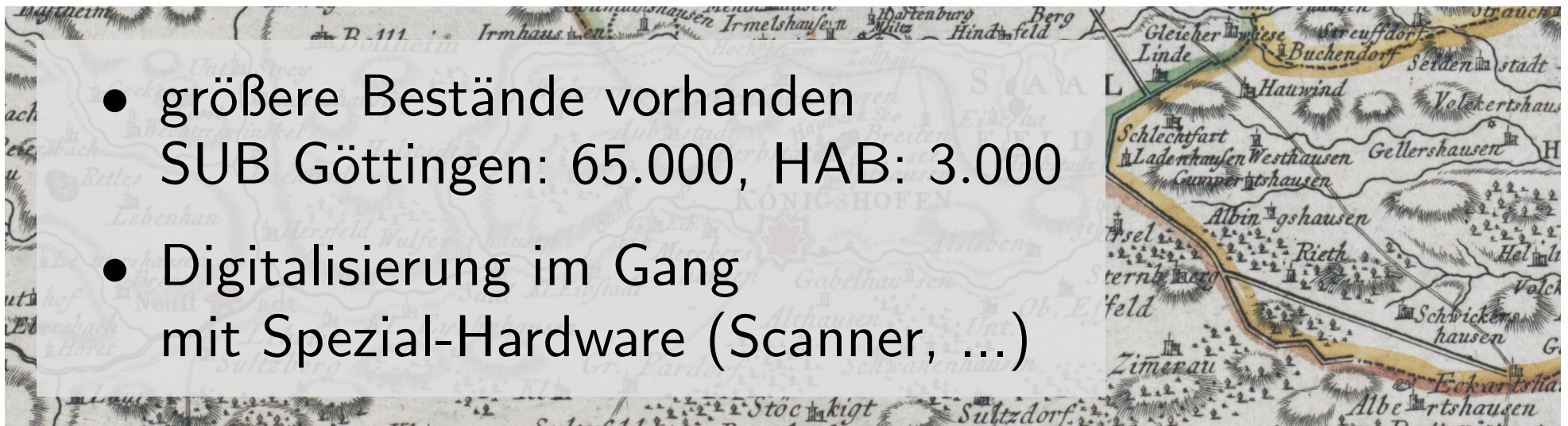
- größere Bestände vorhanden
SUB Göttingen: 65.000, HAB: 3.000
- Digitalisierung im Gang



Einführung

Algorithmen zur Analyse historischer Landkarten

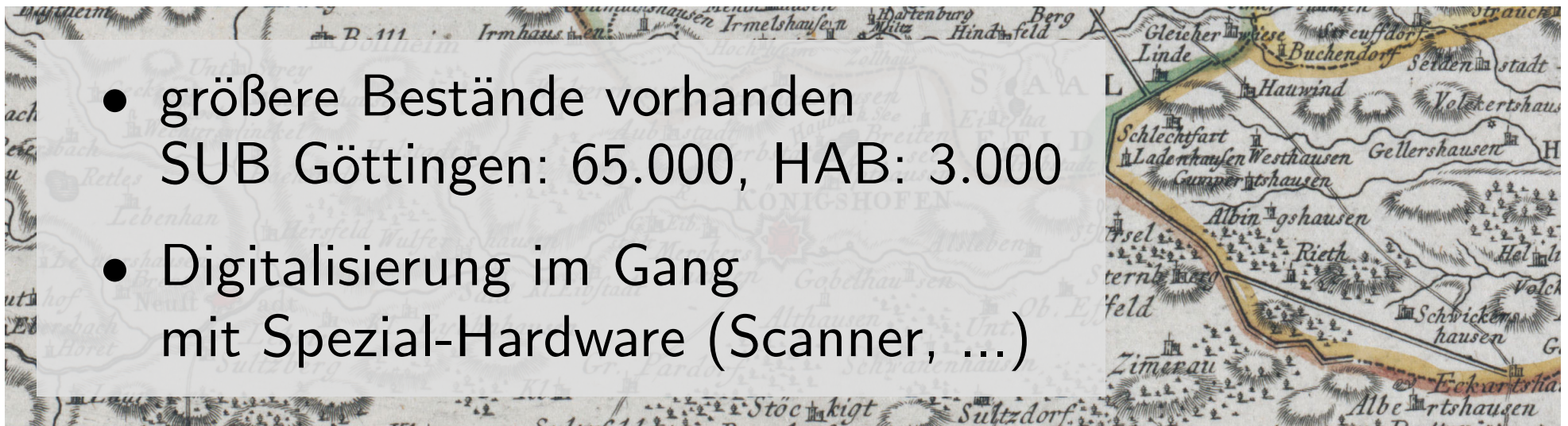
- größere Bestände vorhanden
SUB Göttingen: 65.000, HAB: 3.000
- Digitalisierung im Gang
mit Spezial-Hardware (Scanner, ...)



Einführung

Algorithmen zur Analyse historischer Landkarten

- größere Bestände vorhanden
SUB Göttingen: 65.000, HAB: 3.000
- Digitalisierung im Gang
mit Spezial-Hardware (Scanner, ...)



Einführung

Algorithmen zur Analyse historischer Landkarten



Einführung

- enthaltene Ortschaften

Algorithmen zur Analyse historischer Landkarten



Einführung

- enthaltene Ortschaften
- Geländetopographie

Algorithmen zur Analyse historischer Landkarten



Einführung

- enthaltene Ortschaften
- Geländetopographie
→ Wälder, Flüsse, ...

Algorithmen zur Analyse historischer Landkarten



Einführung

- enthaltene Ortschaften
- Geländetopographie
→ Wälder, Flüsse, ...
- und vieles mehr!

Algorithmen zur Analyse historischer Landkarten



Einführung

- enthalten Ortschaften
- Geländetopographie
- Flüsse, ...
- und vieles mehr!

Algorithmen zur Analyse historischer Landkarten



Einführung

- enthalten Ortschaften
- Geländetopographie
- Flüsse, ...
- und vieles mehr!

Algorithmen zur Analyse historischer Landkarten



Einführung

- enthalten Ortschaften
- Geländetopographie
- Flüsse, ...
- und vieles mehr!

Algorithmen zur Analyse historischer Landkarten



Einführung

Zur Erhebung solcher Meta-
daten sind pro Karte

- enthaltenen Ortschaften
- Geländetopographie
- Flüsse, ...
- und vieles mehr!

Algorithmen zur Analyse historischer Landkarten



Einführung

Zur Erhebung solcher Meta-
daten sind pro Karte
etwa 30 Arbeitsstunden
notwendig.

- enthalten Ortschaften
- Geländetopographie
- Flüsse, ...
- und vieles mehr!

Algorithmen zur Analyse historischer Landkarten



Einführung

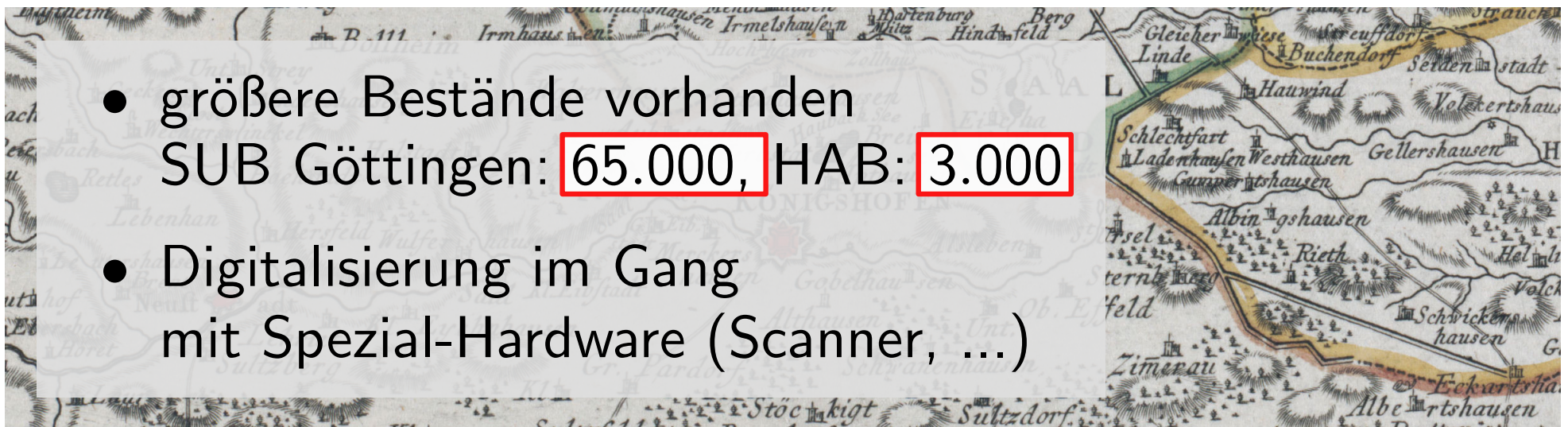
Zur Erhebung solcher Meta-
daten sind pro Karte
etwa 30 Arbeitsstunden
notwendig.

- enthalten Ortschaften
- Geländetopographie
- Wälder, Flüsse, ...
- und vieles mehr!

**wertvolle
Metadaten**

Algorithmen zur Analyse historischer Landkarten

- größere Bestände vorhanden
SUB Göttingen: **65.000**, HAB: **3.000**
- Digitalisierung im Gang
mit Spezial-Hardware (Scanner, ...)



Elemente historischer Landkarten

Elemente historischer Landkarten

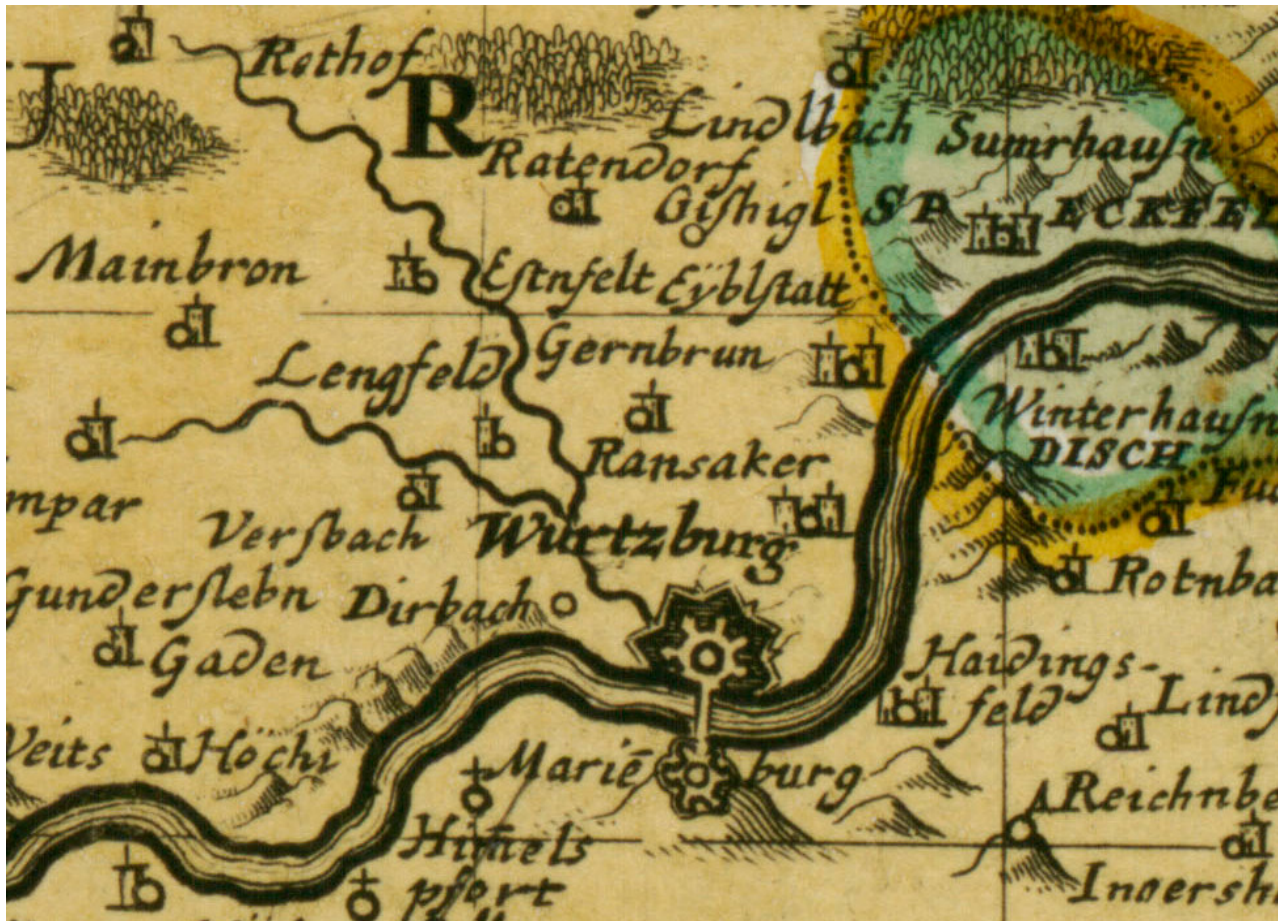
Karten-Quelle für diesen Vortrag:

Franconica-Sammlung der Uni-Bibliothek Würzburg

Elemente historischer Landkarten

Karten-Quelle für diesen Vortrag:

Franconica-Sammlung der Uni-Bibliothek Würzburg



hergestellt im Jahr 1676

Elemente historischer Landkarten

Karten-Quelle für diesen Vortrag:

Franconica-Sammlung der Uni-Bibliothek Würzburg



← Norden

hergestellt im Jahr 1676

Elemente historischer Landkarten

Karten-Quelle für diesen Vortrag:

Franconica-Sammlung der Uni-Bibliothek Würzburg



Wald

← Norden

Elemente historischer Landkarten

Karten-Quelle für diesen Vortrag:

Franconica-Sammlung der Uni-Bibliothek Würzburg



Wald

Fluss

← Norden

Elemente historischer Landkarten

Karten-Quelle für diesen Vortrag:

Franconica-Sammlung der Uni-Bibliothek Würzburg



Wald

Fluss

Erhebung

← Norden

Elemente historischer Landkarten

Karten-Quelle für diesen Vortrag:

Franconica-Sammlung der Uni-Bibliothek Würzburg



Wald

Fluss

Erhebung

**Gelände-
topogr.**

Elemente historischer Landkarten

Karten-Quelle für diesen Vortrag:

Franconica-Sammlung der Uni-Bibliothek Würzburg

Label



Wald

Fluss

Erhebung

Gelände-
topogr.

Elemente historischer Landkarten

Karten-Quelle für diesen Vortrag:

Franconica-Sammlung der Uni-Bibliothek Würzburg

Orts-
markierung

Label



Wald

Fluss

Erhebung

**Gelände-
topogr.**

Elemente historischer Landkarten

Karten-Quelle für diesen Vortrag:

Franconica-Sammlung der Uni-Bibliothek Würzburg

Orts-
markierung

Label

Siedlungs-
topogr.



Wald

Fluss

Erhebung

Gelände-
topogr.

← Norden

Elemente historischer Landkarten

Extraktionsstrategie:

Orts-
markierung
Label
Siedlungs-
topogr.



Wald
Fluss
Erhebung
Gelände-
topogr.

Elemente historischer Landkarten

Extraktionsstrategie: finde die relevanten Elemente...



Elemente historischer Landkarten

Extraktionsstrategie: finde die relevanten Elemente...
... und lies die zugehörigen Beschriftungen



Verarbeitungs-Pipeline

Verarbeitungs-Pipeline

Input

Verarbeitungs-Pipeline

Input



Historische
Landkarte

Verarbeitungs-Pipeline

Input



Historische
Landkarte

Output



Georeferen-
zierte Karte

Verarbeitungs-Pipeline

Input



Historische
Landkarte

Output



Georeferen-
zierte Karte



Index enth.
Ortschaften

Verarbeitungs-Pipeline

Input



Historische
Landkarte

Output



Georeferen-
zierte Karte



Index enth.
Ortschaften



Gelände-
merkmale

Verarbeitungs-Pipeline

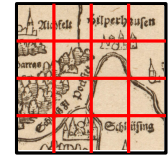
Input



Historische
Landkarte

Segmentierung,
feature extraction

Output



Georeferen-
zierte Karte



Index enth.
Ortschaften



Gelände-
merkmale

Verarbeitungs-Pipeline

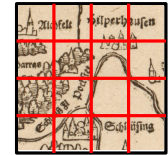
Input



Historische
Landkarte

Segmentierung,
feature extraction

Output



Georeferen-
zierte Karte



Index enth.
Ortschaften



Gelände-
merkmale

Verarbeitungs-Pipeline

Input

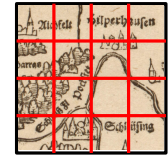


Historische
Landkarte

Segmentierung,
feature extraction

Feedback

Output



Georeferen-
zierte Karte



Index enth.
Ortschaften



Gelände-
merkmale

Verarbeitungs-Pipeline

Input



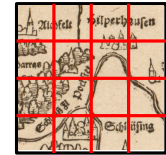
Historische
Landkarte

Segmentierung,
feature extraction

Analyse Gelände-
topographie

Feedback

Output



Georeferen-
zierte Karte



Index enth.
Ortschaften



Gelände-
merkmale

Verarbeitungs-Pipeline

Input



Historische
Landkarte

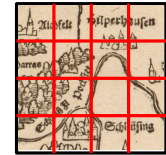
Segmentierung,
feature extraction

Feedback

Analyse Gelände-
topographie

Zuordnung der
Beschriftungen

Output



Georeferen-
zierte Karte



Index enth.
Ortschaften



Gelände-
merkmale

Verarbeitungs-Pipeline

Input



Historische
Landkarte

Segmentierung,
feature extraction

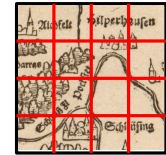
Feedback

Analyse Gelände-
topographie

Zuordnung der
Beschriftungen

OCR von Be-
schriftungen

Output



Georeferen-
zierte Karte



Index enth.
Ortschaften



Gelände-
merkmale

Verarbeitungs-Pipeline

Input



Historische
Landkarte



Wörterbuch mit
hist. Ortsnamen

Segmentierung,
feature extraction

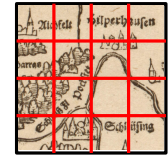
Feedback

Analyse Gelände-
topographie

Zuordnung der
Beschriftungen

OCR von Be-
schriftungen

Output



Georeferen-
zierte Karte

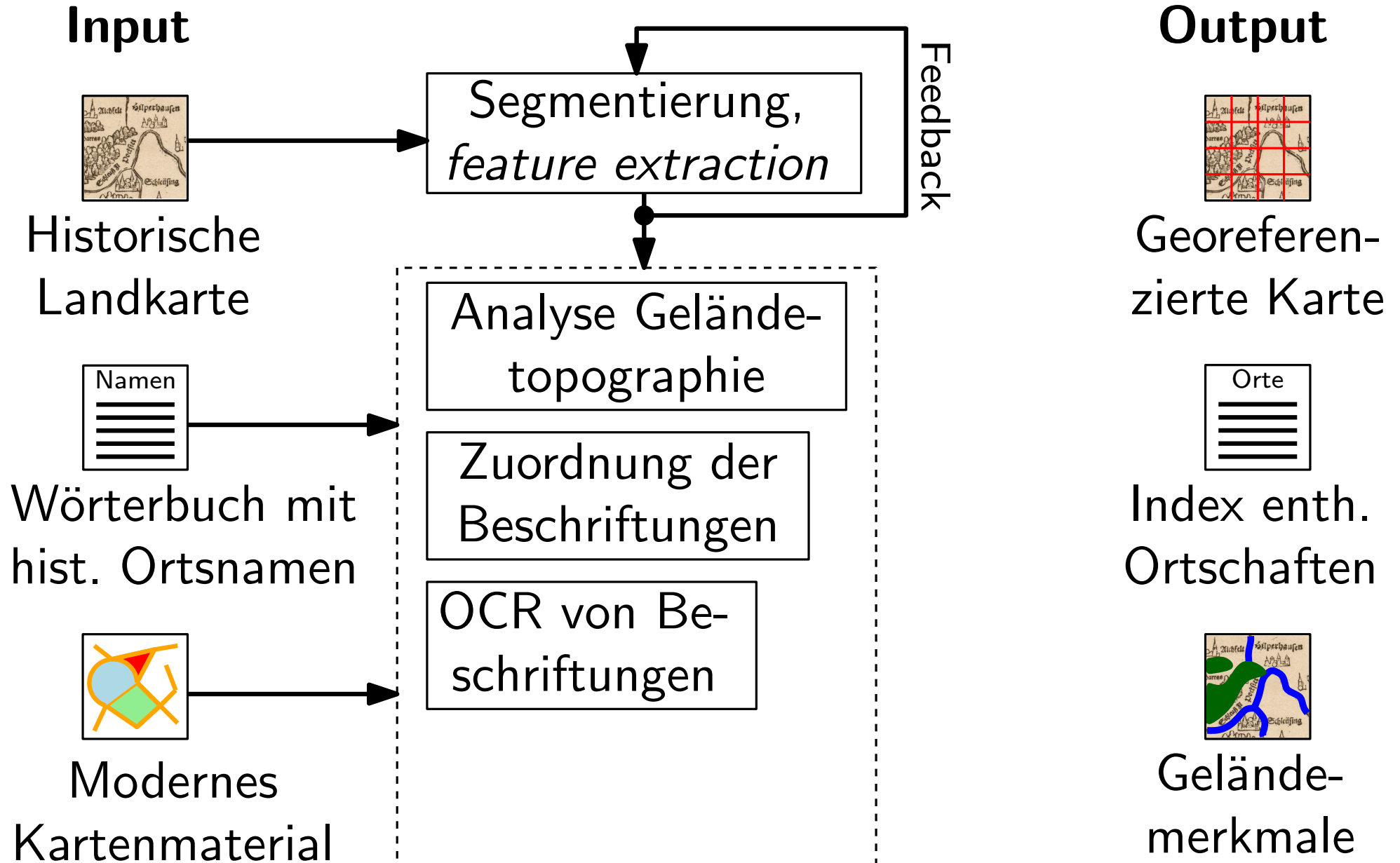


Index enth.
Ortschaften

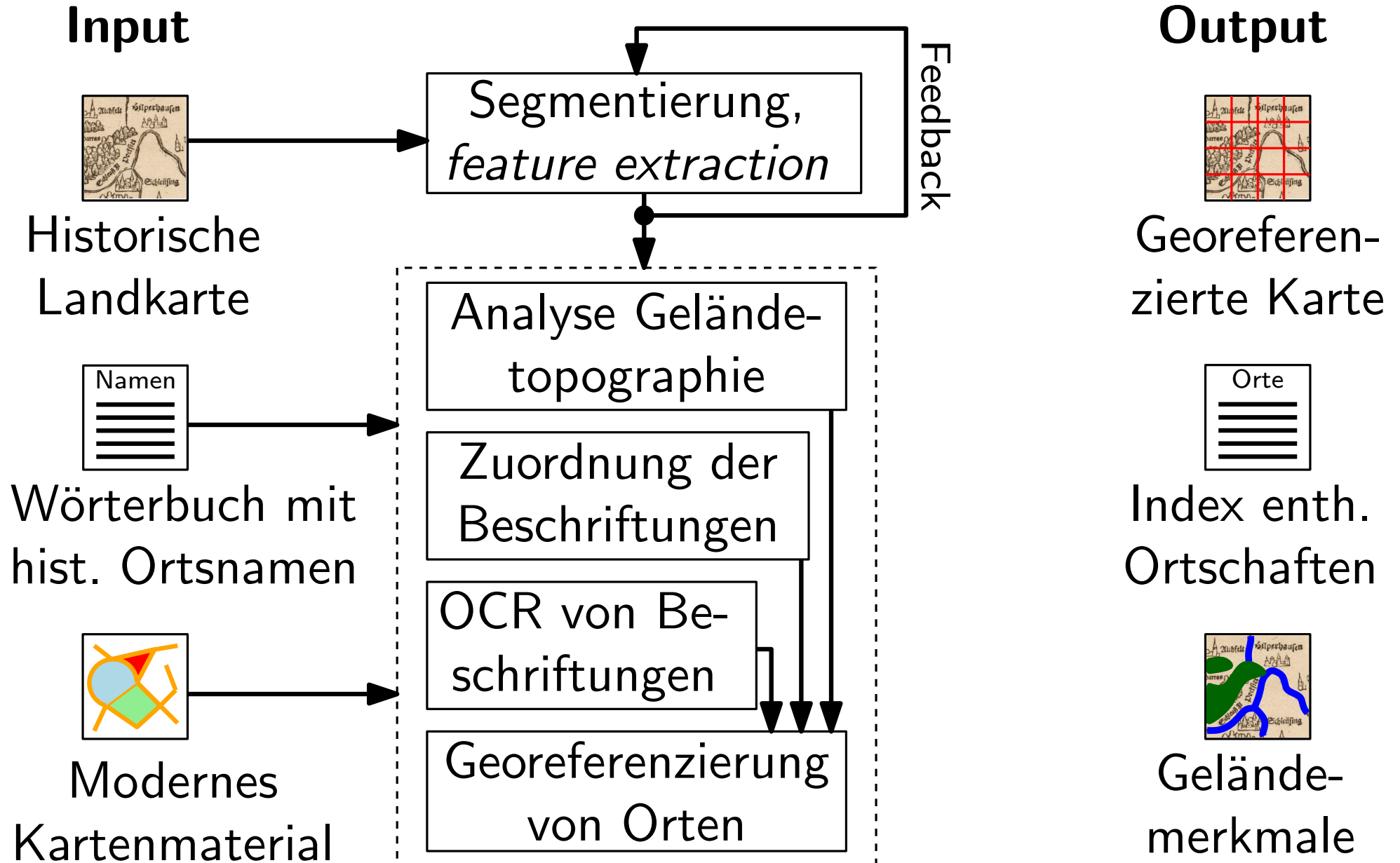


Gelände-
merkmale

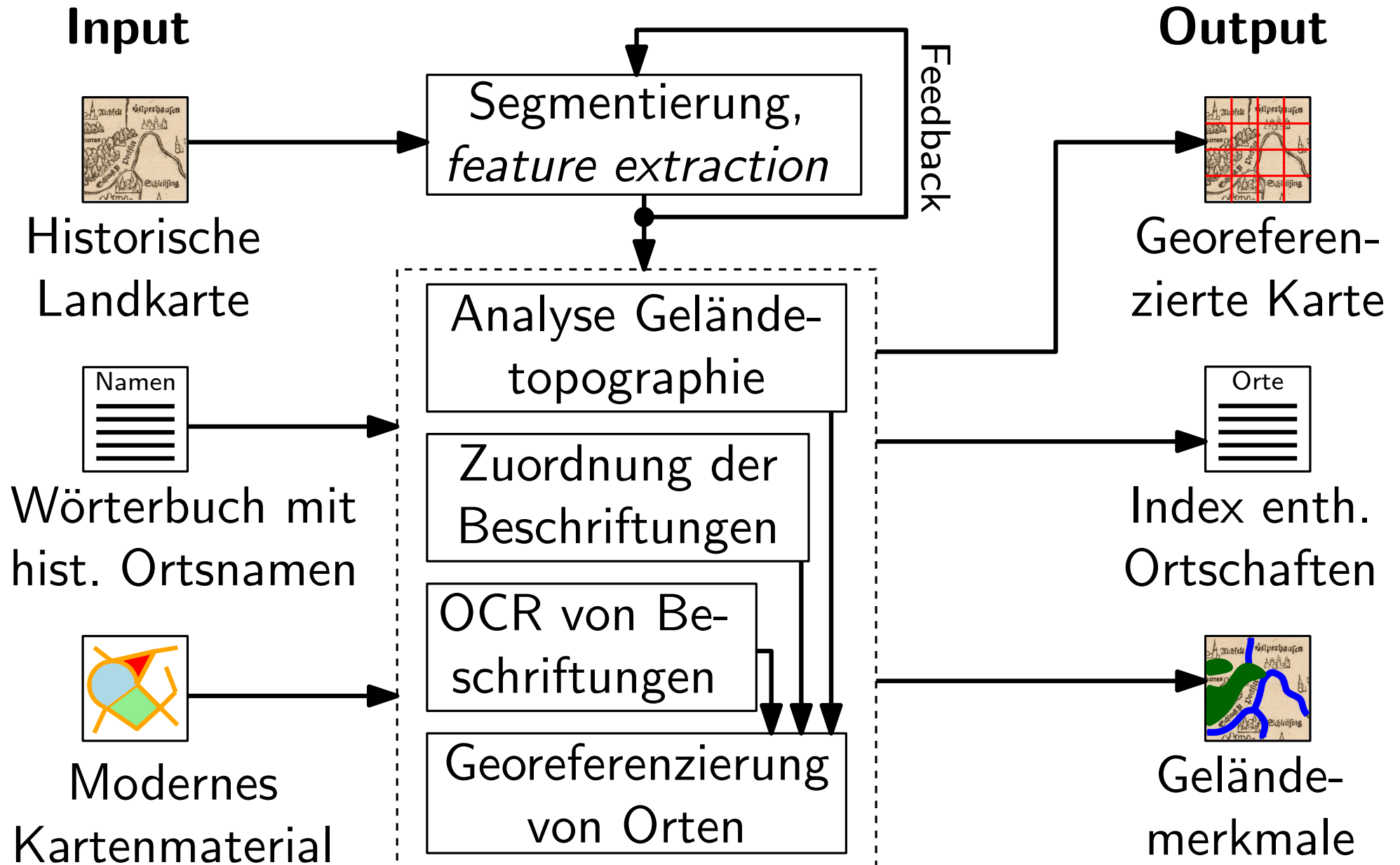
Verarbeitungs-Pipeline



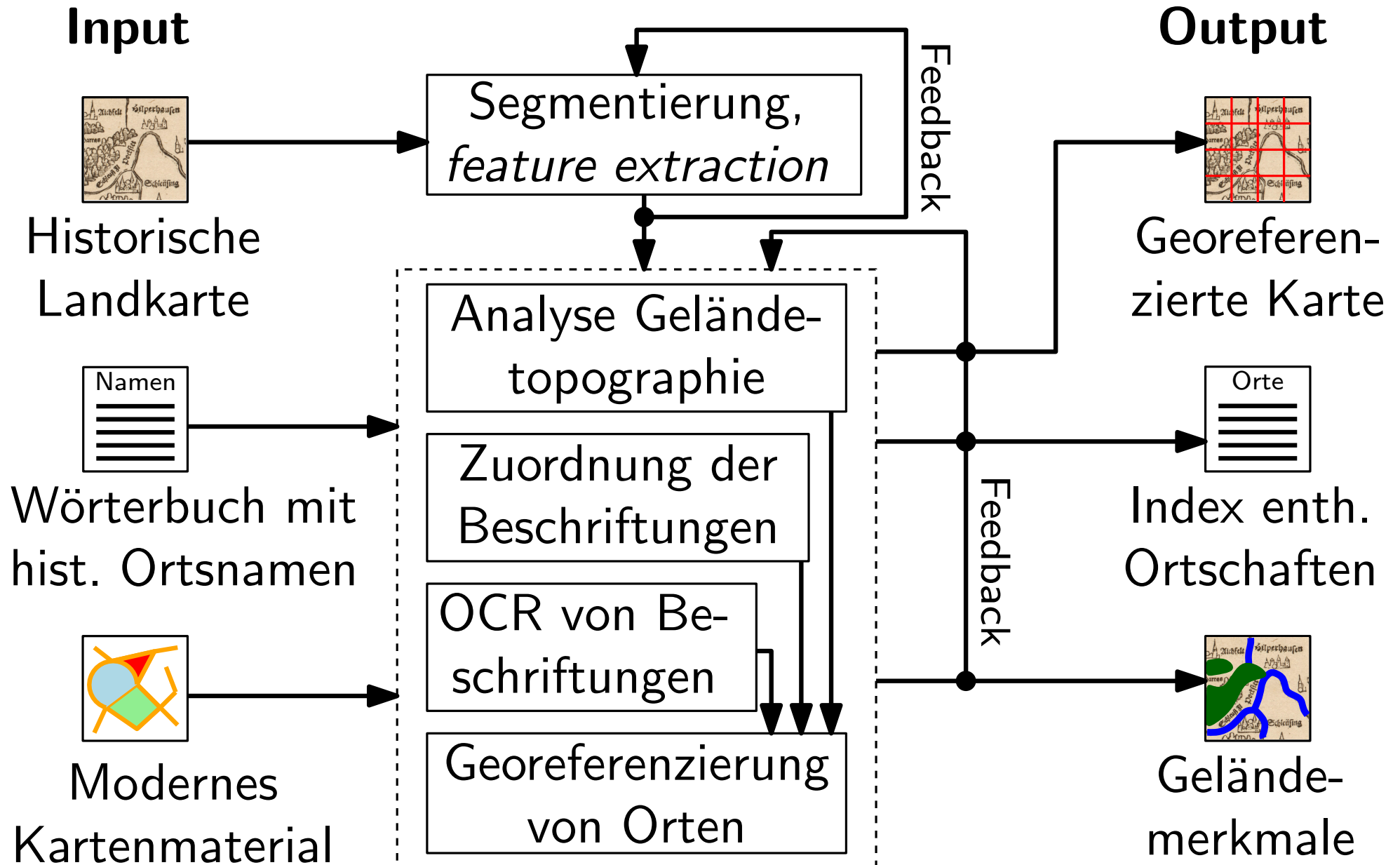
Verarbeitungs-Pipeline



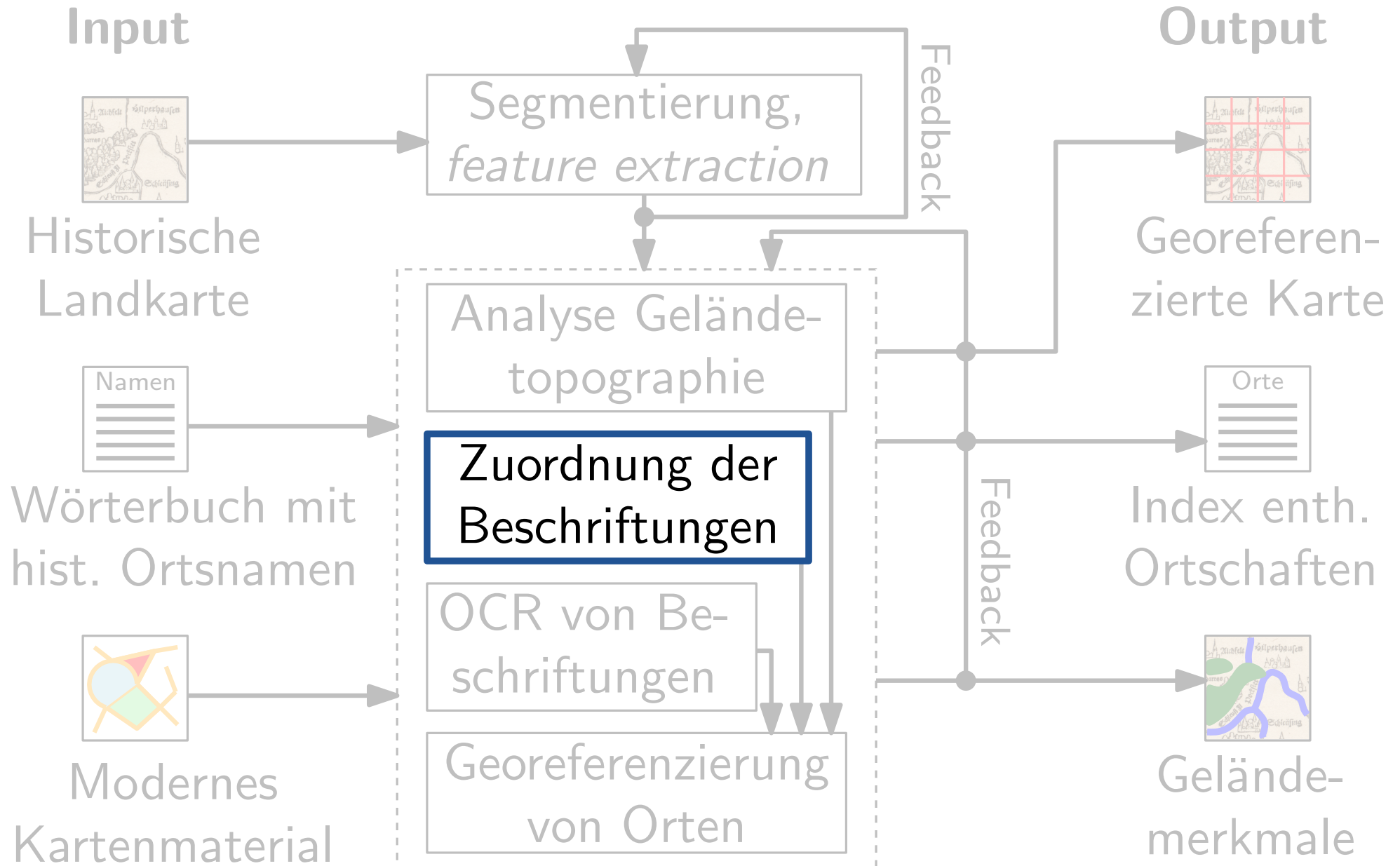
Verarbeitungs-Pipeline



Verarbeitungs-Pipeline



Verarbeitungs-Pipeline



Ausgangssituation

Ausgangssituation

Input:

Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P , dargestellt durch achsenparallele *bounding boxes*.

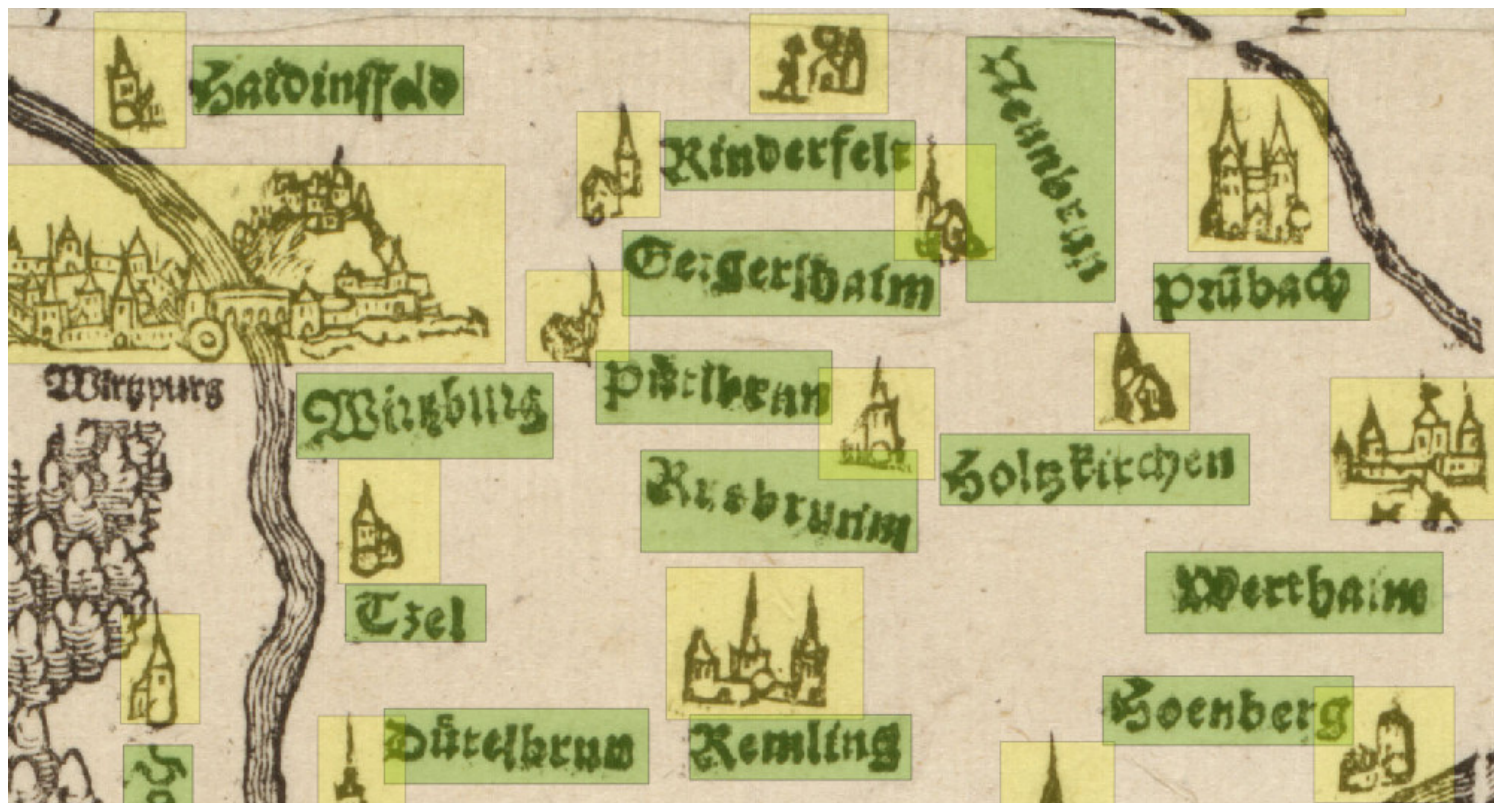
Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P , dargestellt durch achsenparallele *bounding boxes*.



Ausgangssituation

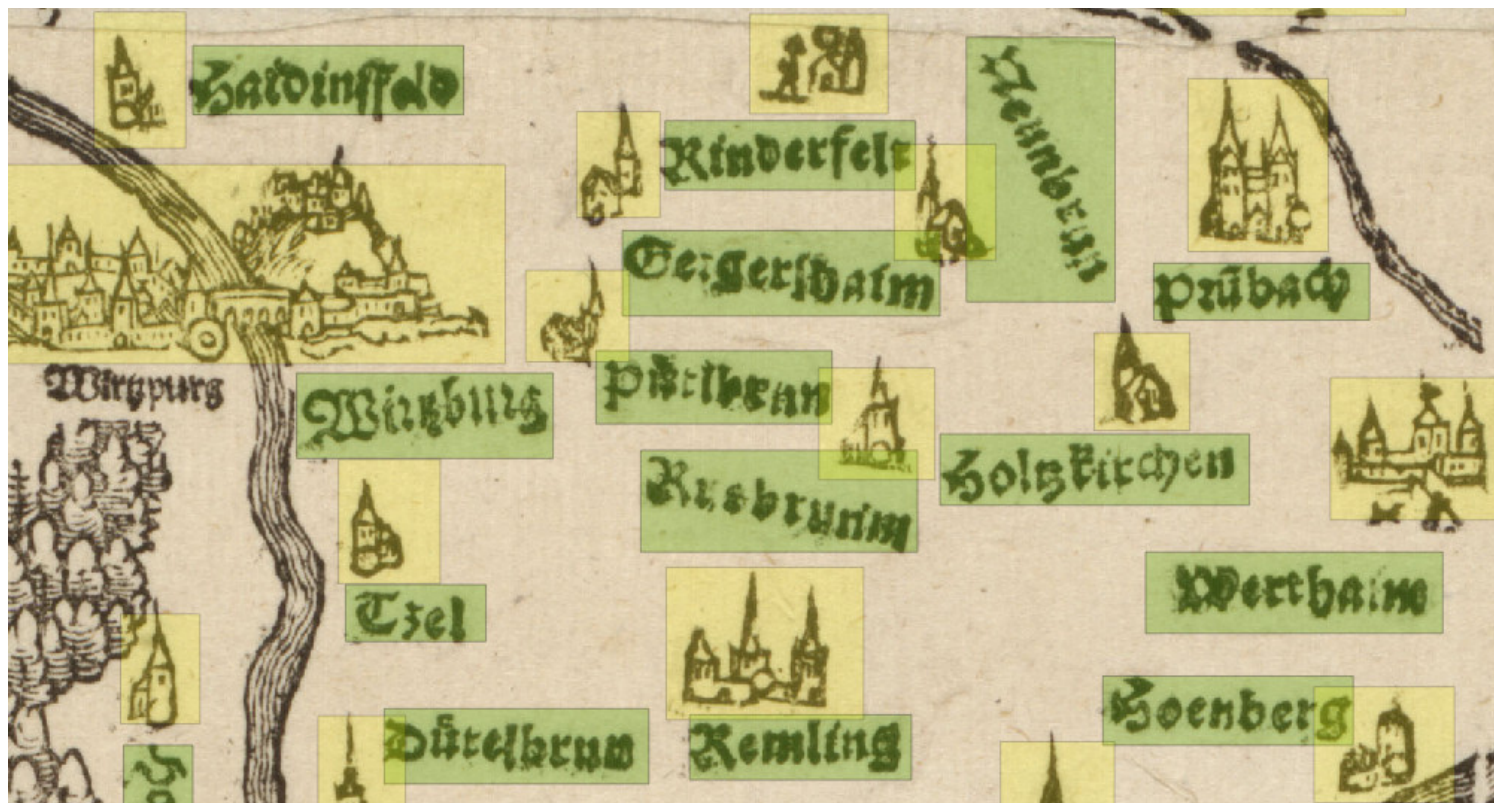
- Input:**
- Menge von Ortsmarkierungen P , dargestellt durch achsenparallele *bounding boxes*.
 - Menge von Labels L , ebenfalls dargestellt durch achsenparallele *bounding boxes*.



Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P
 - Menge von Labels L

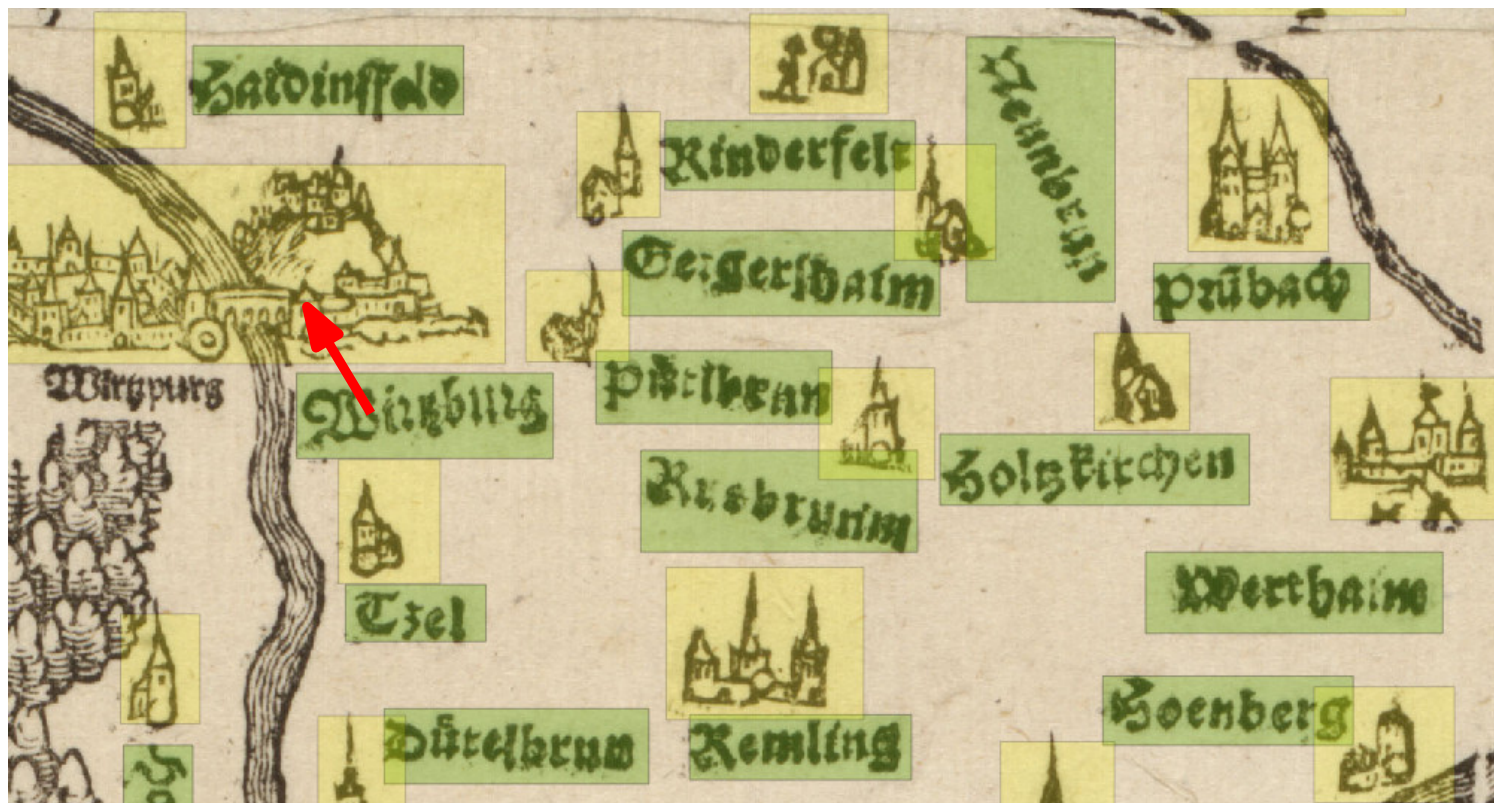
Problem: Welches Label beschriftet welche Ortsmarkierung?



Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P
 - Menge von Labels L

Problem: Welches Label beschriftet welche Ortsmarkierung?



Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P
 - Menge von Labels L

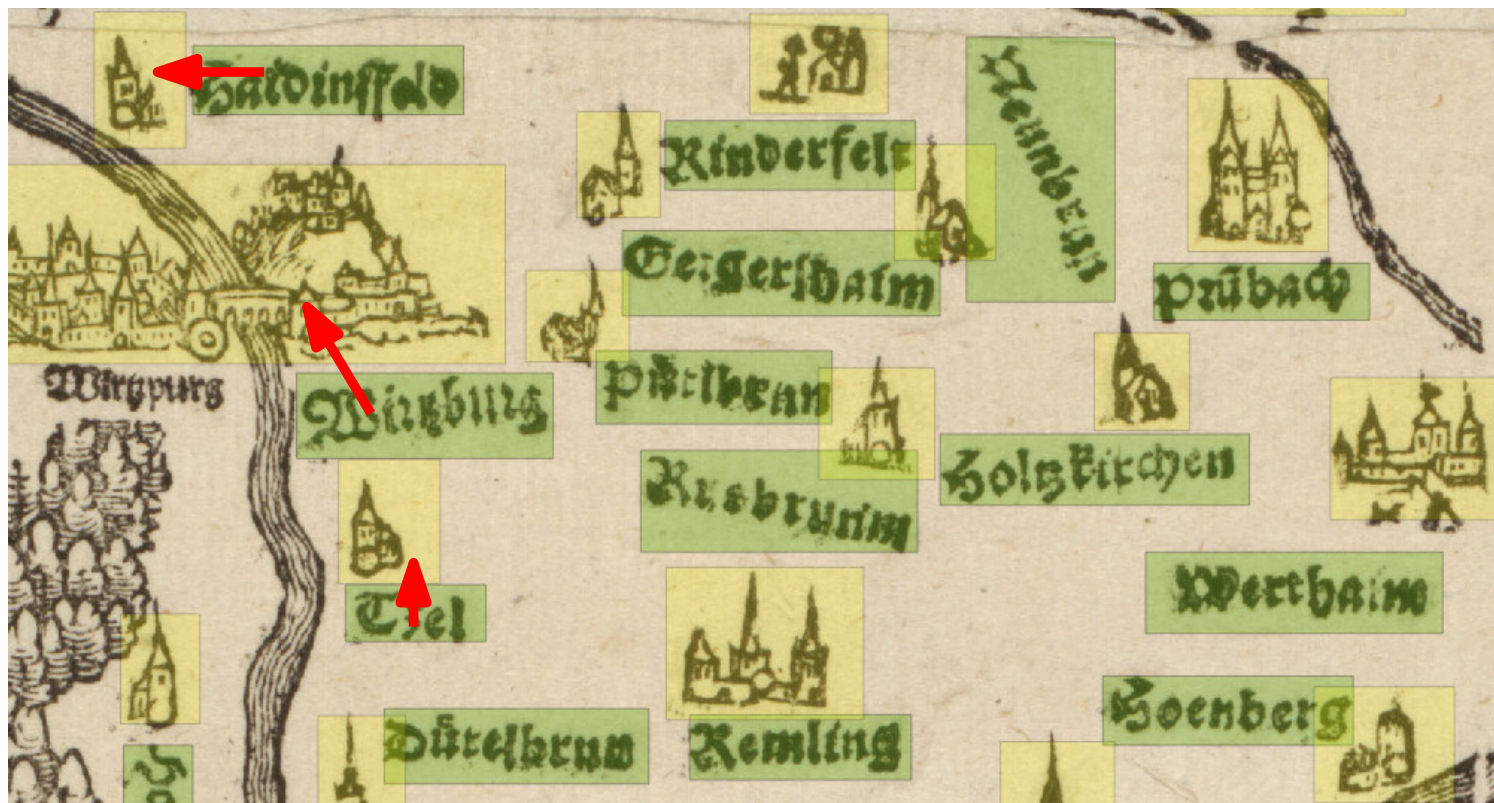
Problem: Welches Label beschriftet welche Ortsmarkierung?



Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P
 - Menge von Labels L

Problem: Welches Label beschriftet welche Ortsmarkierung?



Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P
 - Menge von Labels L

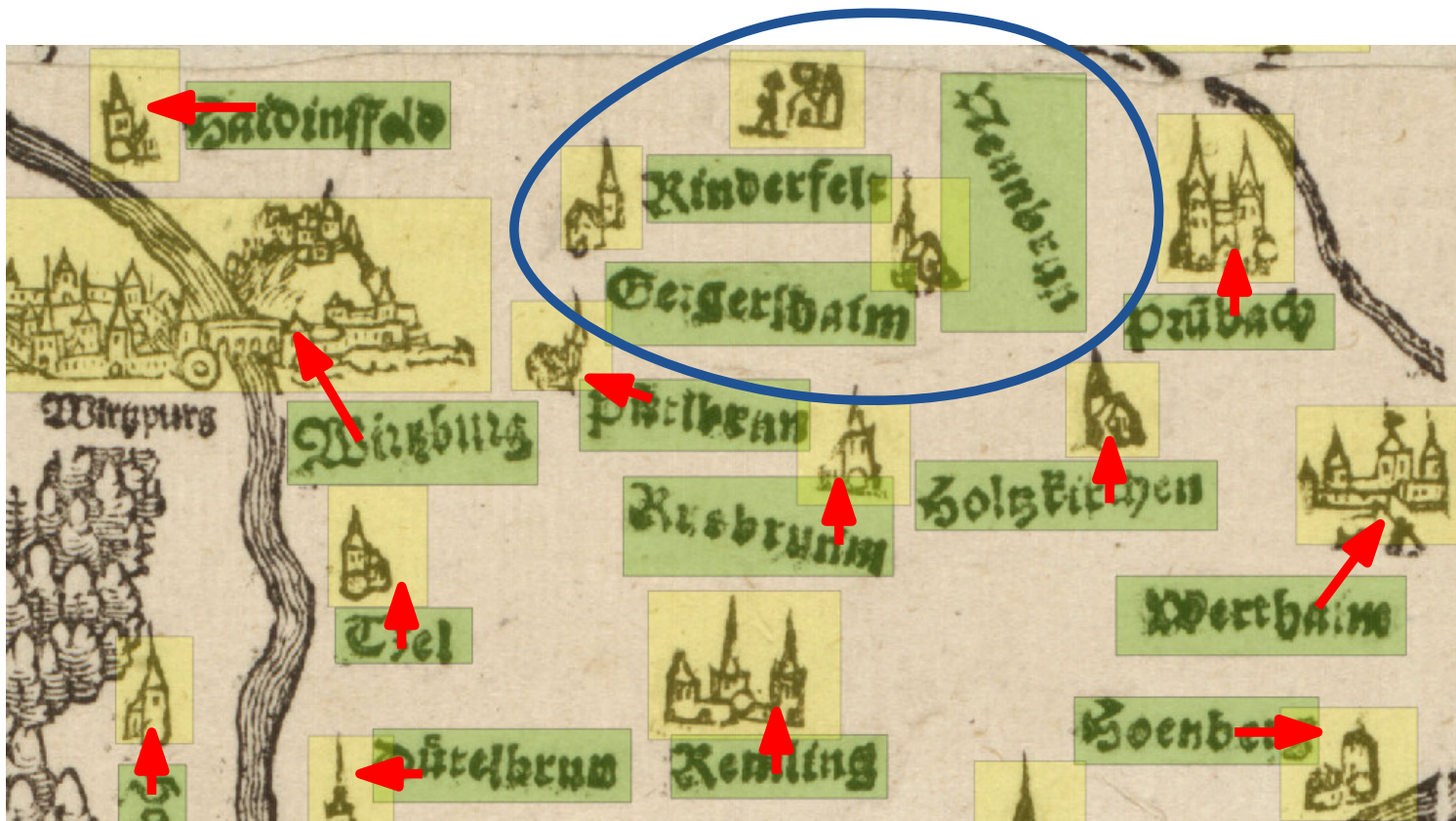
Problem: Welches Label beschriftet welche Ortsmarkierung?



Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P
 - Menge von Labels L

Problem: Welches Label beschriftet welche Ortsmarkierung?



Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P
 - Menge von Labels L

Problem: Welches Label beschriftet welche Ortsmarkierung?



Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P
 - Menge von Labels L

Problem: Welches Label beschriftet welche Ortsmarkierung?



Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P
 - Menge von Labels L

Problem: Welches Label beschriftet welche Ortsmarkierung?



Ausgangssituation

- Input:**
- Menge von Ortsmarkierungen P
 - Menge von Labels L

Problem: Welches Label beschriftet welche Ortsmarkierung?



Dieses Problem ist schwierig, auch für Menschen!

Modellierung als Optimierungsproblem

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass

- jedes $p \in P$ höchstens einem $l \in L$ und
- jedes $l \in L$ höchstens einem $p \in P$

zugeordnet ist.

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass

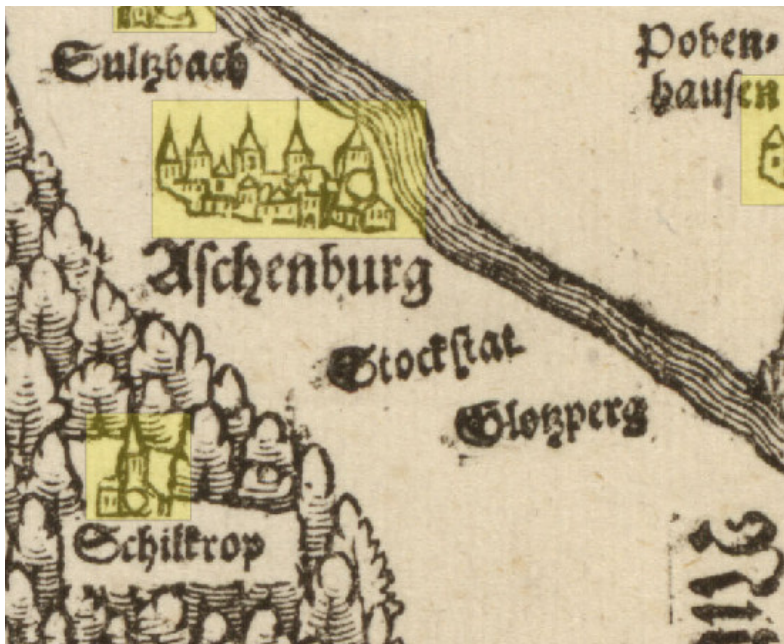
- jedes $p \in P$ höchstens einem $l \in L$ und
- jedes $l \in L$ höchstens einem $p \in P$

zugeordnet ist.

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass

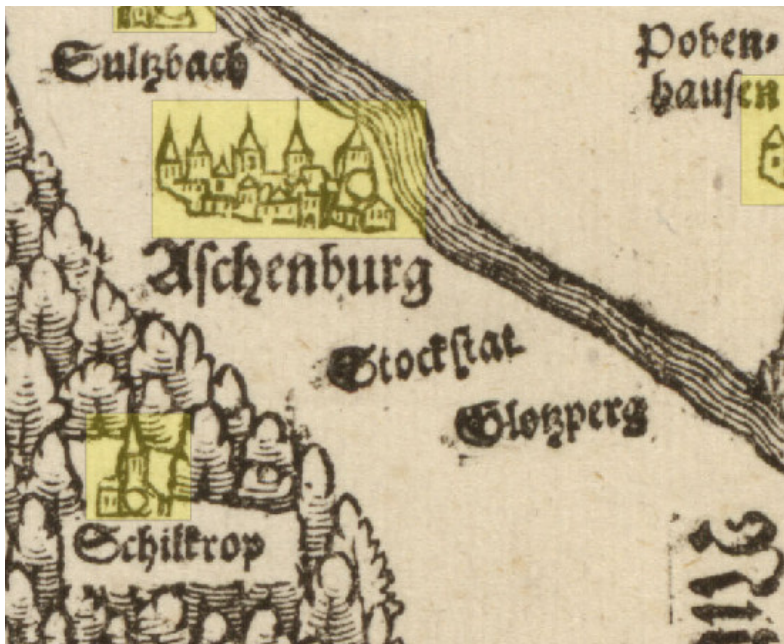
- jedes $p \in P$ höchstens einem $l \in L$ und
- jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.



Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass

- jedes $p \in P$ höchstens einem $l \in L$ und
- jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

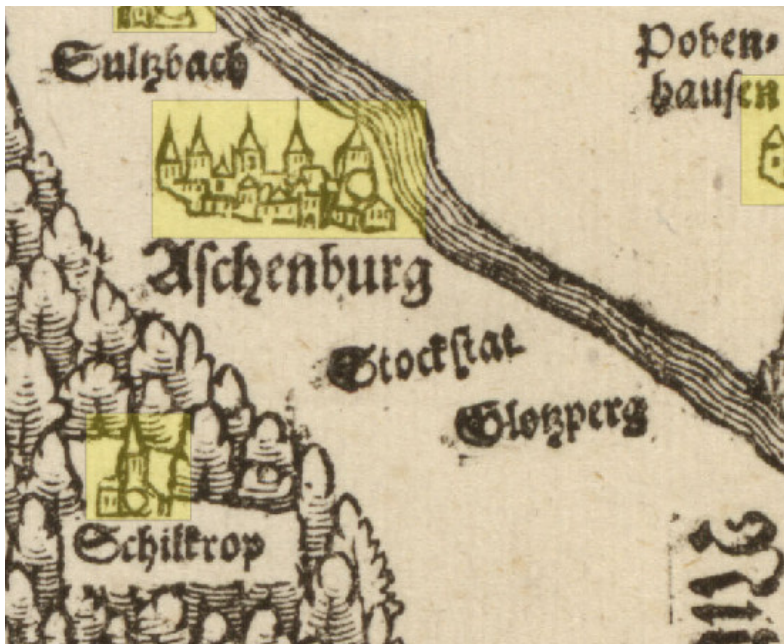


Keine 1-zu-1-Zuordnung aufgrund

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass

- jedes $p \in P$ höchstens einem $l \in L$ und
- jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.



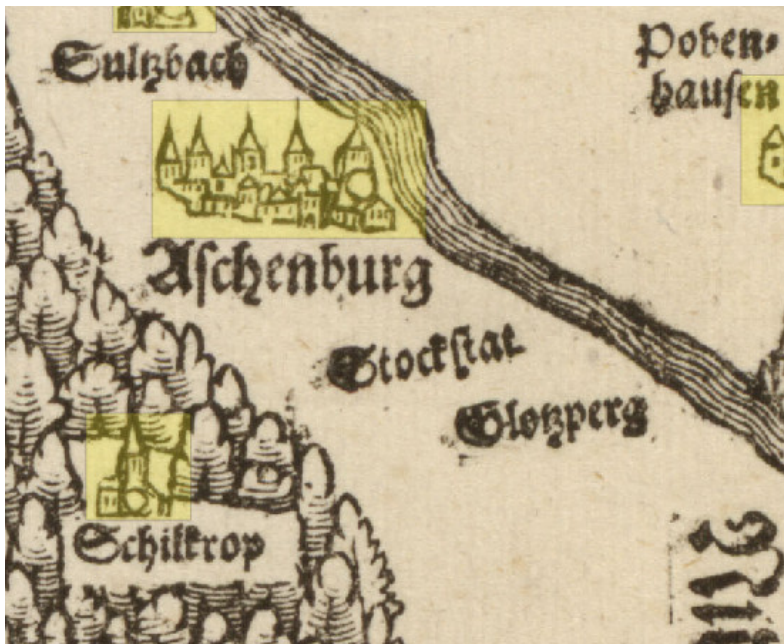
Keine 1-zu-1-Zuordnung aufgrund

- fehlender Elemente in den historischen Karten

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass

- jedes $p \in P$ höchstens einem $l \in L$ und
- jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.



Keine 1-zu-1-Zuordnung aufgrund

- fehlender Elemente in den historischen Karten
- etwaiger Fehler in vorhergehender Segmentierung

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele:

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

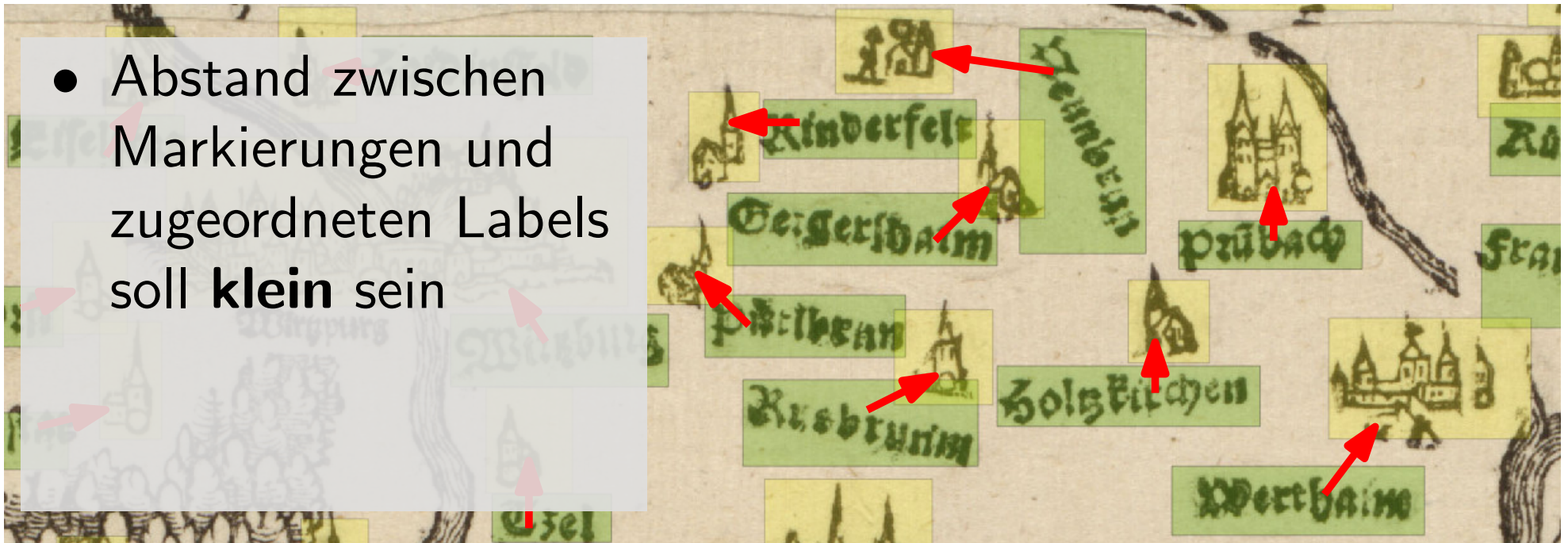


Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- Abstand zwischen Markierungen und zugeordneten Labels soll **klein** sein

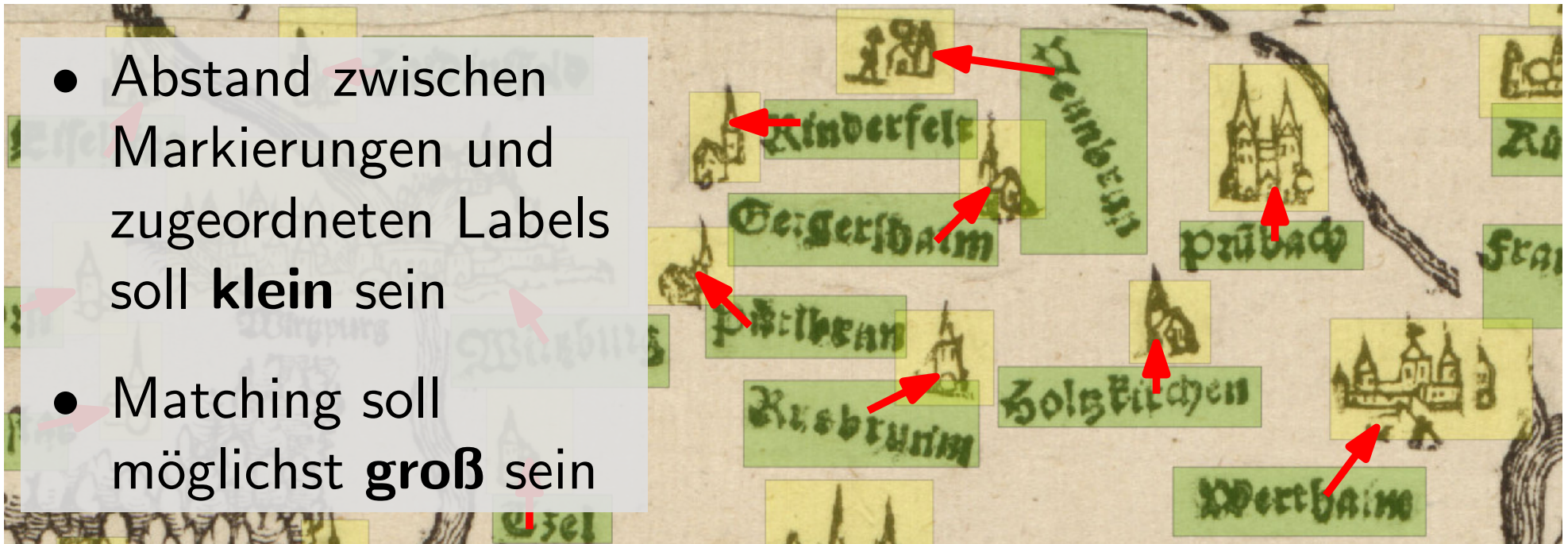


Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- Abstand zwischen Markierungen und zugeordneten Labels soll **klein** sein
- Matching soll möglichst **groß** sein



Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- Abstand zwischen Markierungen und zugeordneten Labels soll **klein** sein
- Matching soll möglichst **groß** sein

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- Abstand zwischen Markierungen und zugeordneten Labels soll **klein** sein
- Matching soll möglichst **groß** sein

(1) Die Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- Abstand zwischen Markierungen und zugeordneten Labels soll **klein** sein
- Matching soll möglichst **groß** sein

- (1) Die Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- Abstand zwischen Markierungen und zugeordneten Labels soll **klein** sein
- Matching soll möglichst **groß** sein

- (1) Die Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente
- (3) Keine Zuordnung $(p, l) \in M$ hat Abstand $d(p, l) > r$

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- Abstand zwischen Markierungen und zugeordneten Labels soll **klein** sein
- Matching soll möglichst **groß** sein

- (1) Die Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente
- (3) Keine Zuordnung $(p, l) \in M$ hat Abstand $d(p, l) > \boxed{r}$

„Suchradius“

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- (1) Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente
- (3) Keine Zuordnung $(p, l) \in M$ mit $d(p, l) > r$

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- (1) Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente
- (3) Keine Zuordnung $(p, l) \in M$ mit $d(p, l) > r$

Greedy-Strategie?

„Nimm die beiden Elemente mit dem geringsten Abstand und wiederhole.“

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- (1) Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente
- (3) Keine Zuordnung $(p, l) \in M$ mit $d(p, l) > r$

Greedy-Strategie?

„Nimm die beiden Elemente mit dem geringsten Abstand und wiederhole.“ **Schlechte Ergebnisse!**

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- (1) Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente
- (3) Keine Zuordnung $(p, l) \in M$ mit $d(p, l) > r$

**Ziel-
funktion:**

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- (1) Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente
- (3) Keine Zuordnung $(p, l) \in M$ mit $d(p, l) > r$

Ziel-

funktion: maximiere $f_{\text{obj}}(M) = \sum_{(p,l) \in M} r - d(p, l)$

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- (1) Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente
- (3) Keine Zuordnung $(p, l) \in M$ mit $d(p, l) > r$

Ziel-

funktion:

$$\text{maximiere } f_{\text{obj}}(M) = \sum_{(p,l) \in M} r - d(p, l)$$

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- (1) Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente
- (3) Keine Zuordnung $(p, l) \in M$ mit $d(p, l) > r$

Ziel-

funktion: maximiere $f_{\text{obj}}(M) = \sum_{(p,l) \in M} r - d(p, l)$

Modellierung als Optimierungsproblem

Gesucht: Ein *Matching* von Labels und Ortsmarkierungen, so dass jedes $p \in P$ höchstens einem $l \in L$ und jedes $l \in L$ höchstens einem $p \in P$ zugeordnet ist.

Ziele: Welche Eigenschaften muss ein Matching M erfüllen, um *gut* zu sein?

- (1) Summe über $d(p, l)$ für alle $(p, l) \in M$ ist klein
- (2) M enthält viele Elemente
- (3) Keine Zuordnung $(p, l) \in M$ mit $d(p, l) > r$

**Ziel-
funktion:**

$$\text{maximiere } f_{\text{obj}}(M) = \sum_{(p,l) \in M} r - d(p, l)$$

Polynomialzeit-Algorithmus

Polynomialzeit-Algorithmus

Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p, l))$ maximal ist.

Polynomialzeit-Algorithmus

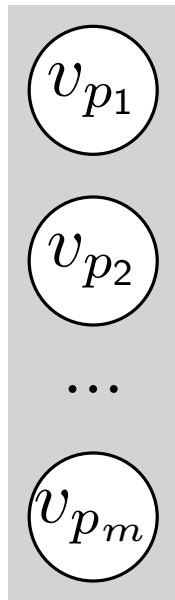
Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p,l))$ maximal ist.

Lösung: Modelliere Problem als Graph $G = (V, E)$...

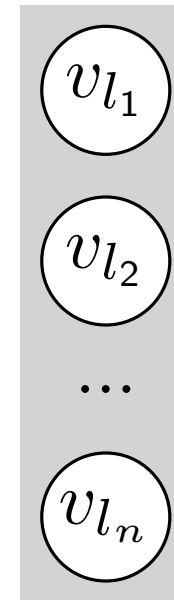
Polynomialzeit-Algorithmus

Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p, l))$ maximal ist.

Lösung: Modelliere Problem als Graph $G = (V, E)$...



Markierungen

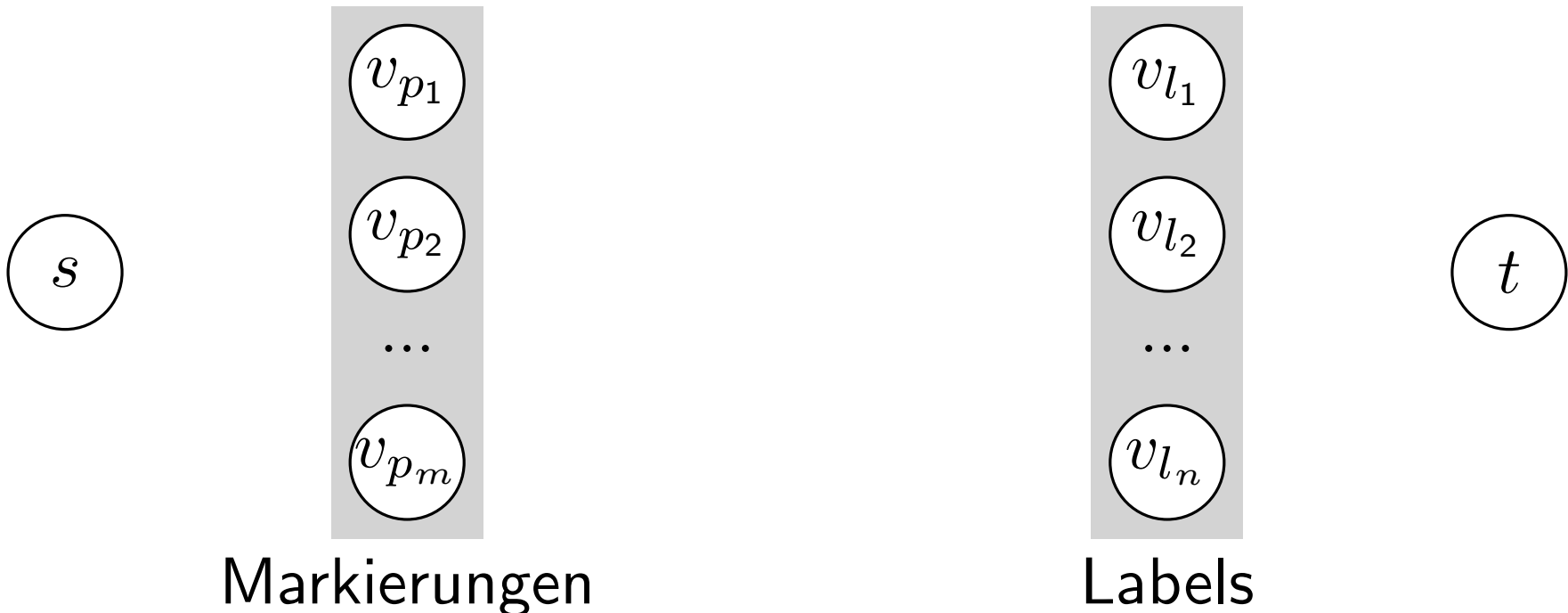


Labels

Polynomialzeit-Algorithmus

Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p, l))$ maximal ist.

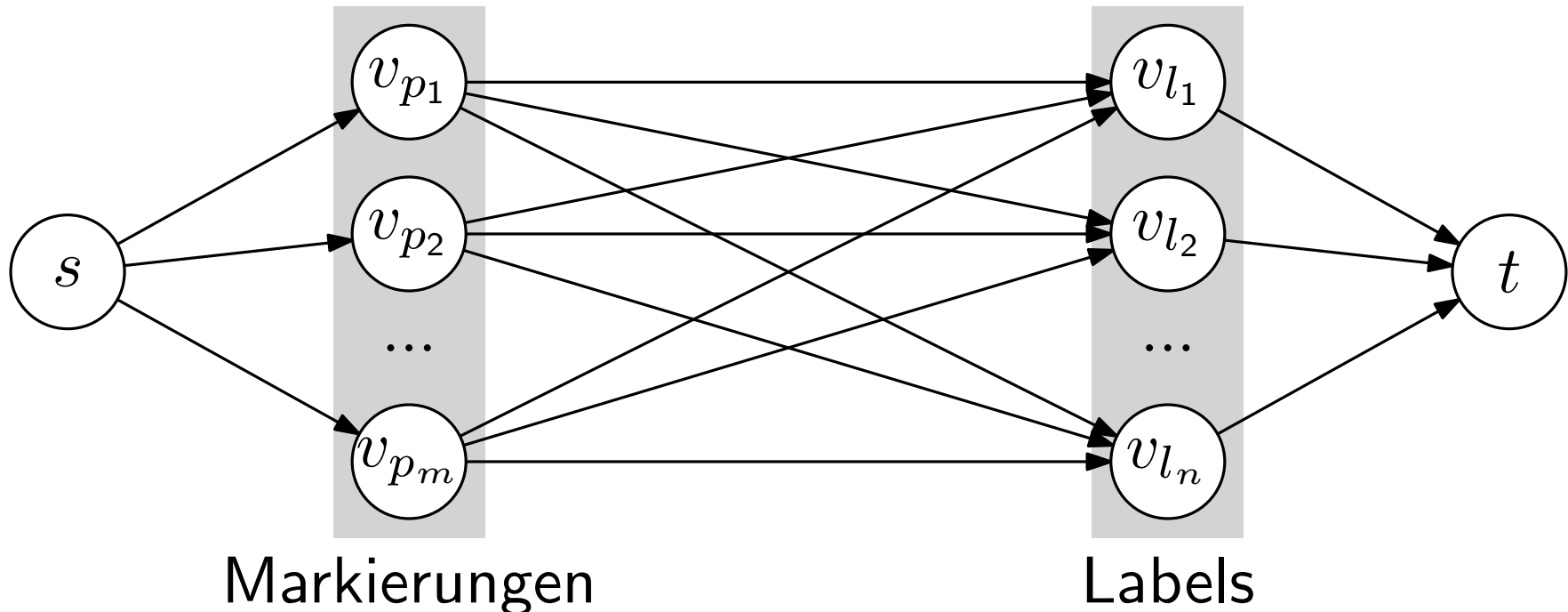
Lösung: Modelliere Problem als Graph $G = (V, E)$...



Polynomialzeit-Algorithmus

Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p, l))$ maximal ist.

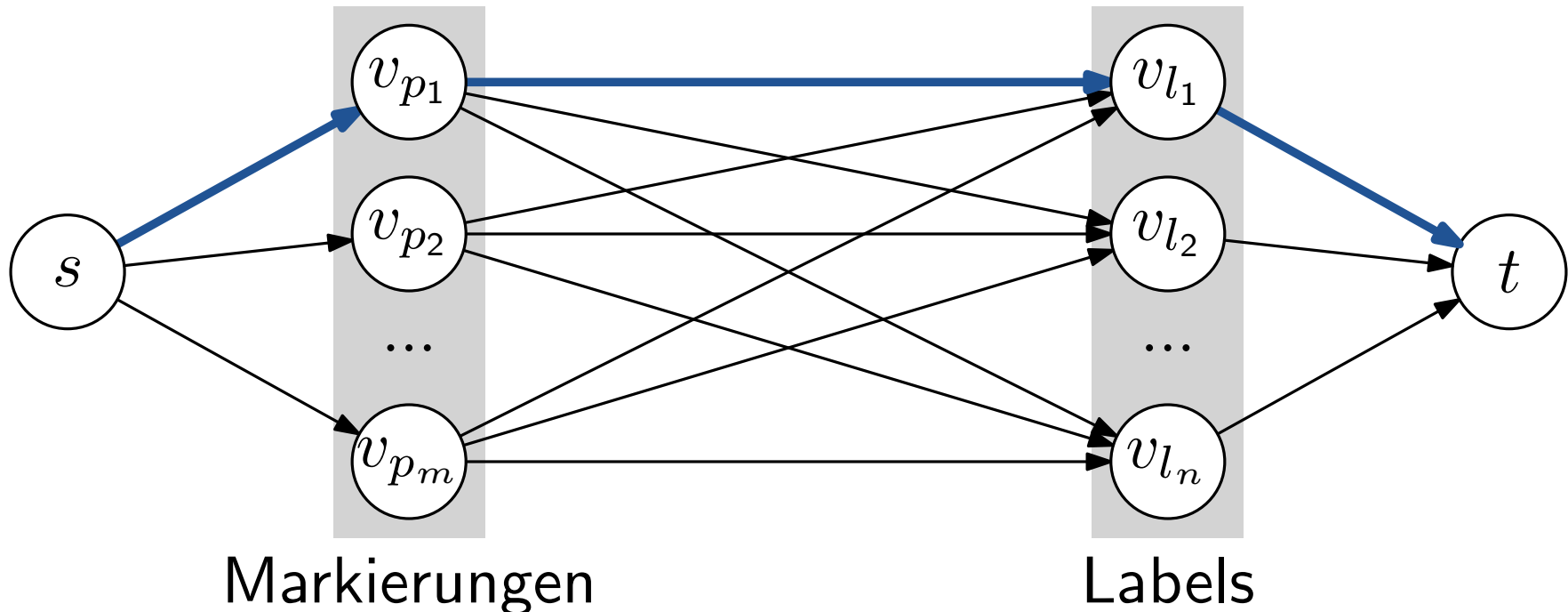
Lösung: Modelliere Problem als Graph $G = (V, E)$...



Polynomialzeit-Algorithmus

Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p, l))$ maximal ist.

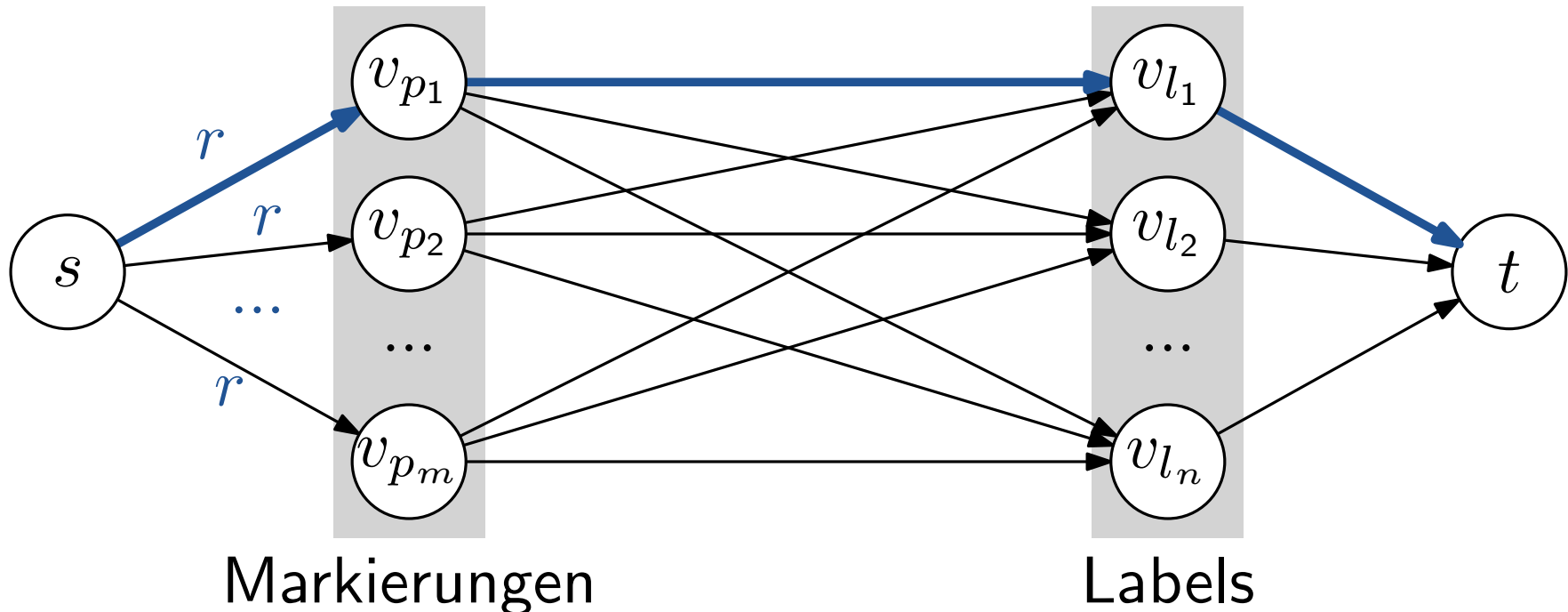
Lösung: Modelliere Problem als Graph $G = (V, E)$...



Polynomialzeit-Algorithmus

Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p,l))$ maximal ist.

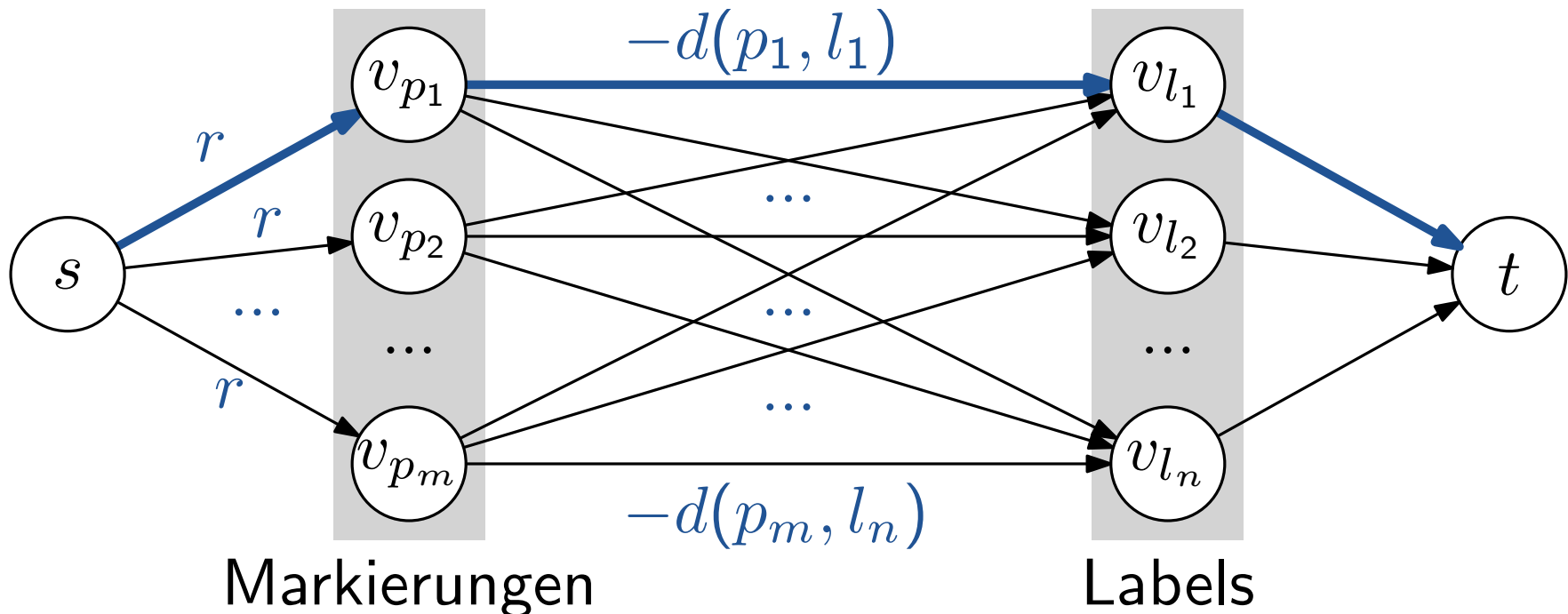
Lösung: Modelliere Problem als Graph $G = (V, E)$...



Polynomialzeit-Algorithmus

Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p,l))$ maximal ist.

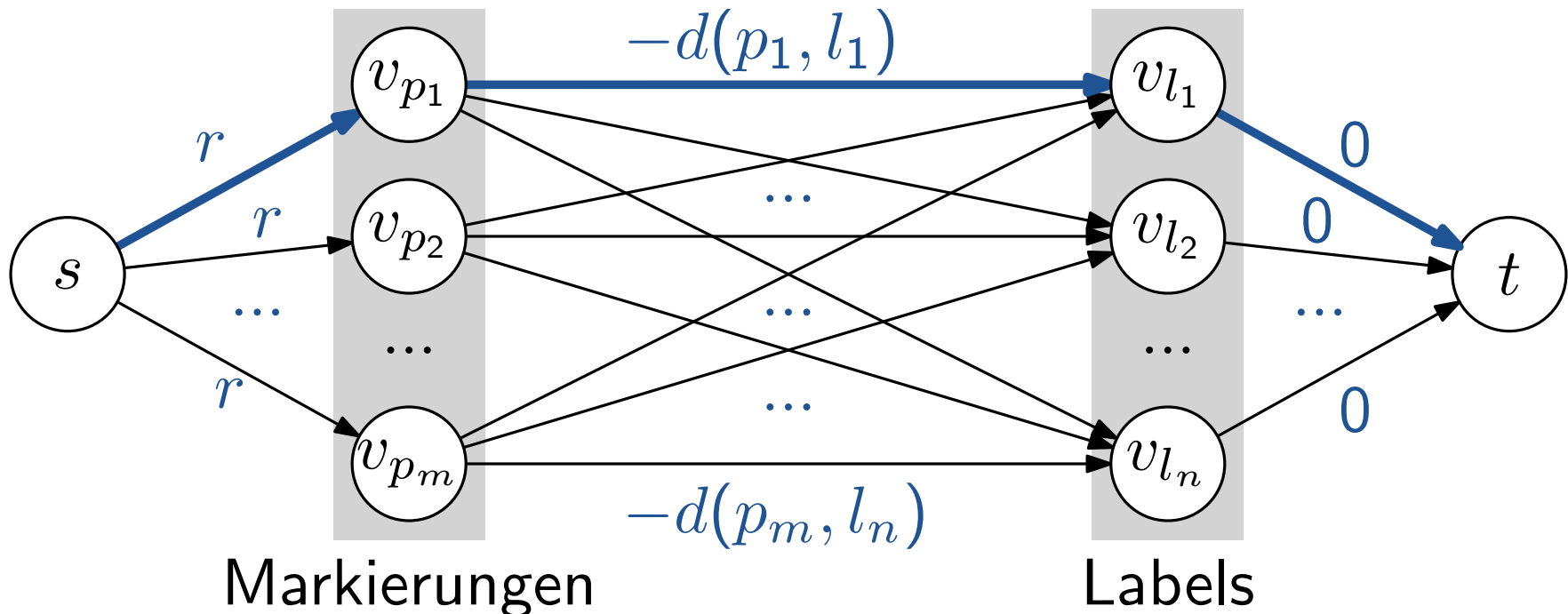
Lösung: Modelliere Problem als Graph $G = (V, E)$...



Polynomialzeit-Algorithmus

Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p,l))$ maximal ist.

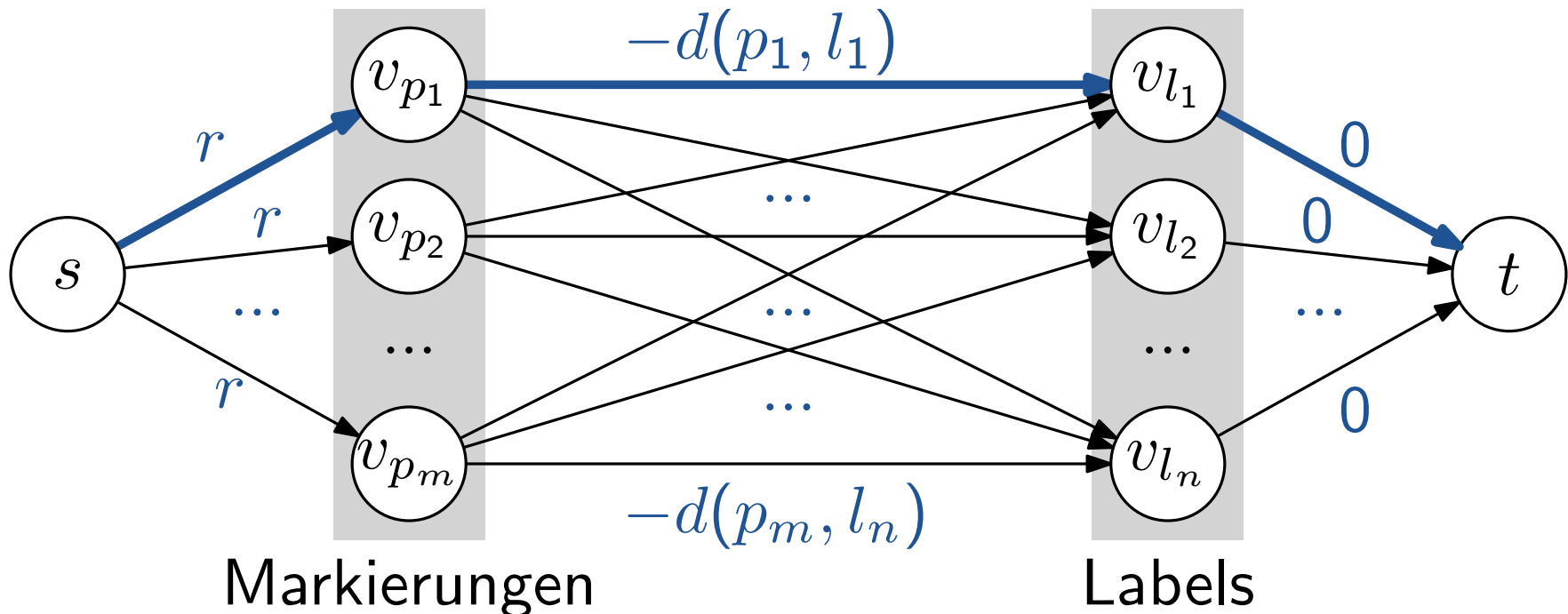
Lösung: Modelliere Problem als Graph $G = (V, E)$...



Polynomialzeit-Algorithmus

Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p,l))$ maximal ist.

Lösung: Modelliere Problem als Graph $G = (V, E)$...
...und finde Fluss mit minimalen Kosten.

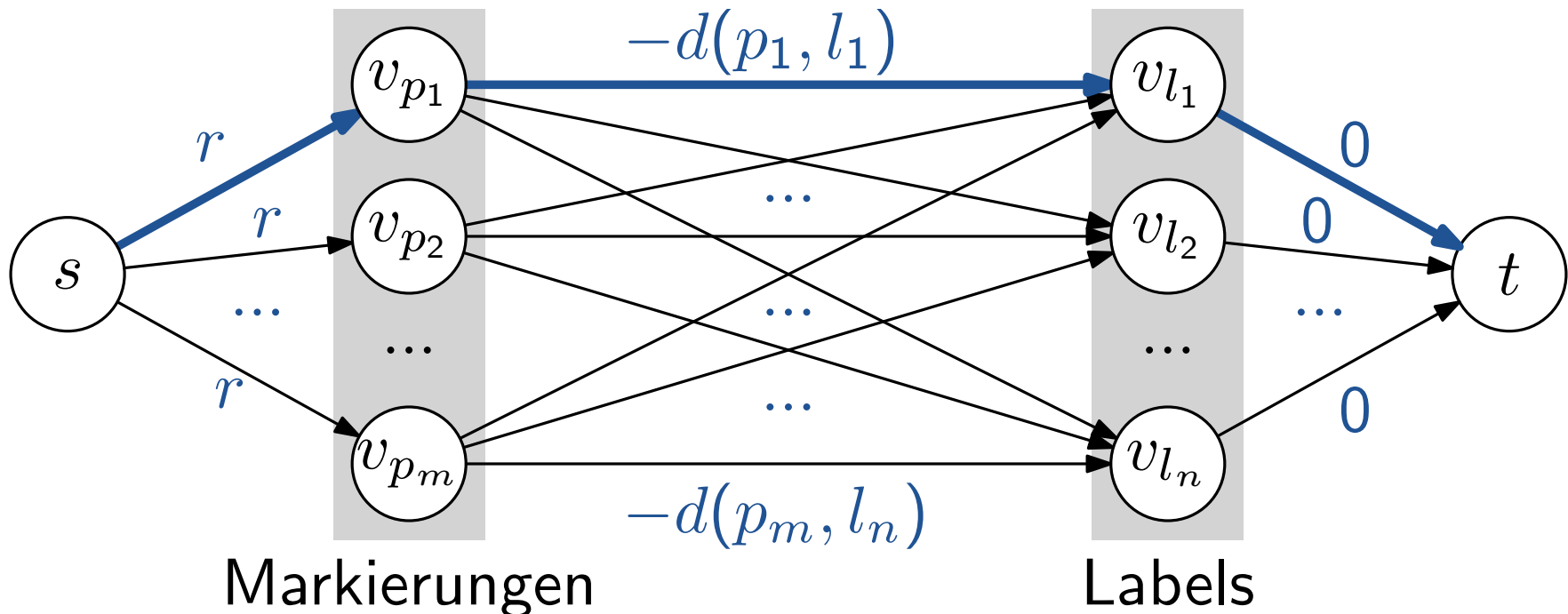


Polynomialzeit-Algorithmus

Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p,l))$ maximal ist.

Lösung: Modelliere Problem als Graph $G = (V, E)$...
...und finde Fluss mit minimalen Kosten.

← effiziente Algorithmen



Polynomialzeit-Algorithmus

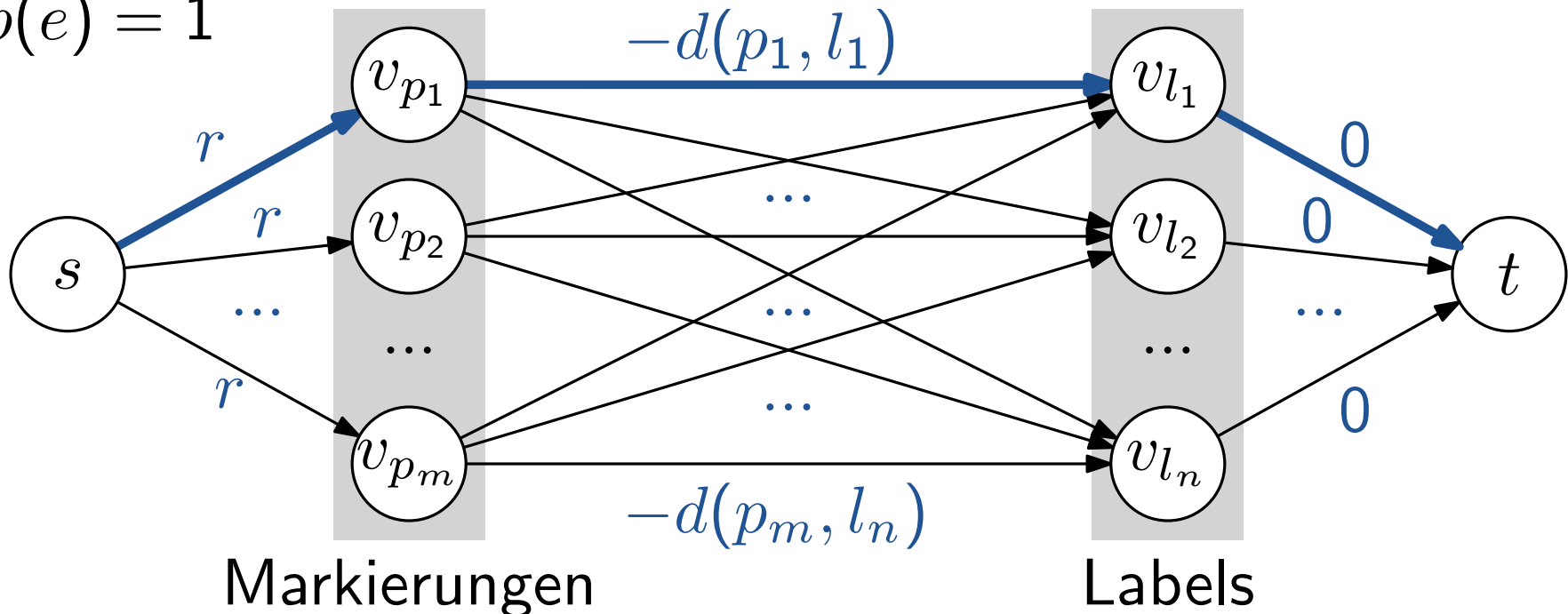
Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p,l))$ maximal ist.

Lösung: Modelliere Problem als Graph $G = (V, E)$...
...und finde Fluss mit minimalen Kosten.

$\forall e \in E:$

$cap(e) = 1$

← effiziente Algorithmen



Polynomialzeit-Algorithmus

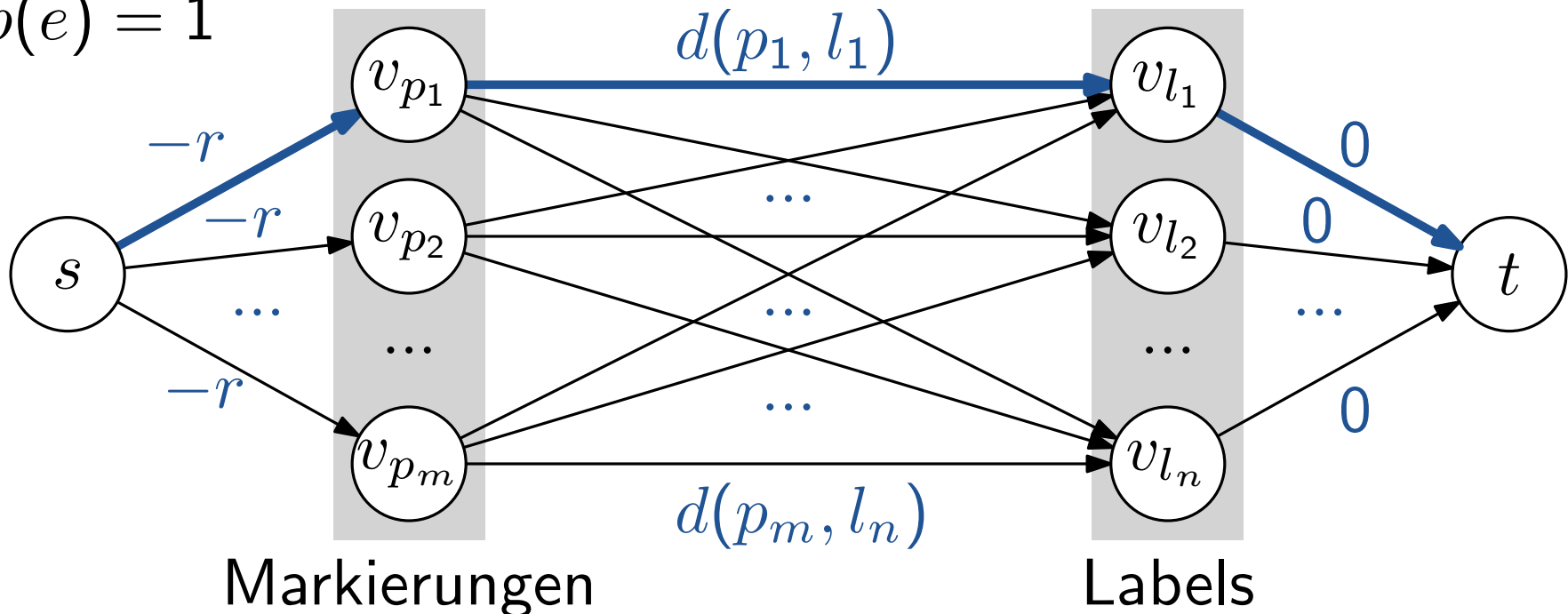
Problem: Finde Matching M von Labels und Markierungen, so dass $f_{\text{obj}}(M) = \sum_{(p,l) \in M} (r - d(p,l))$ maximal ist.

Lösung: Modelliere Problem als Graph $G = (V, E)$...
...und finde Fluss mit minimalen Kosten.

$\forall e \in E:$

$cap(e) = 1$

← effiziente Algorithmen



Experimente

Experimente

Karten:

Experimente

Karten:

Sebastian von Rotenhan:
Das FranckenLandt (1533)



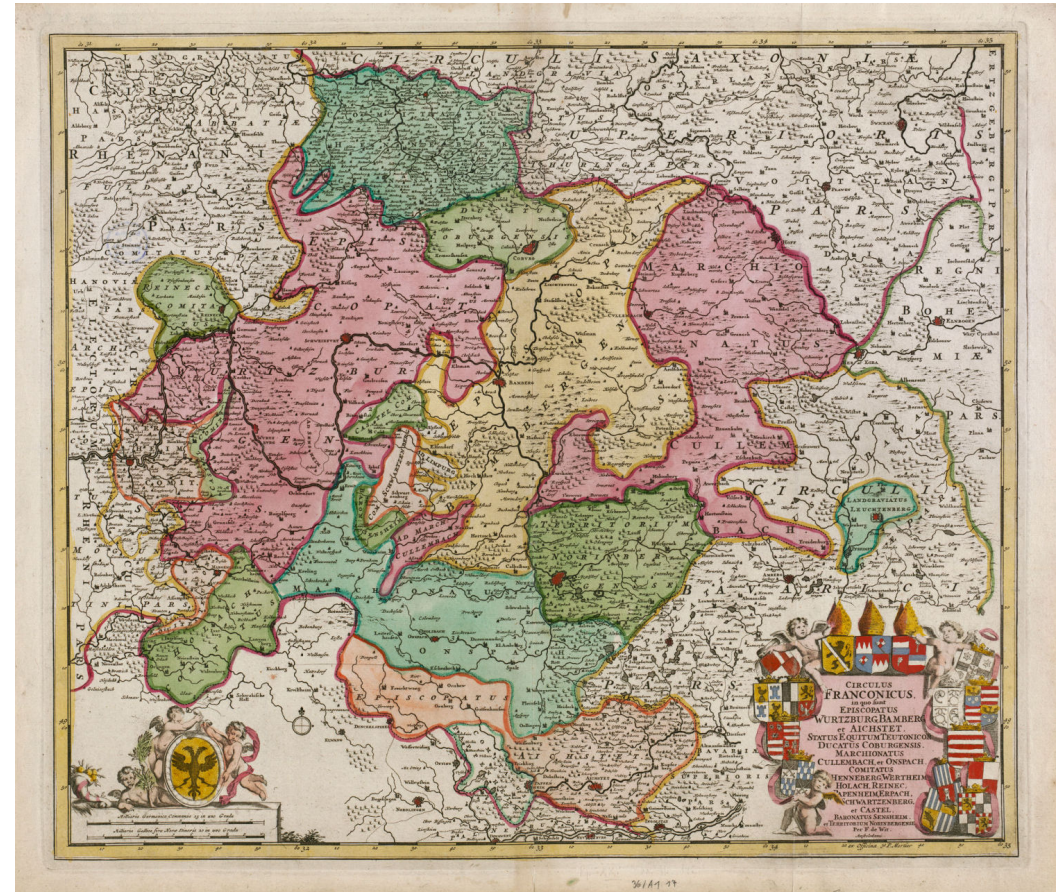
Experimente

Karten:

Sebastian von Rotenhan:
Das FranckenLandt (1533)



Frederik De Wit:
Circulus Franconicus (1706)



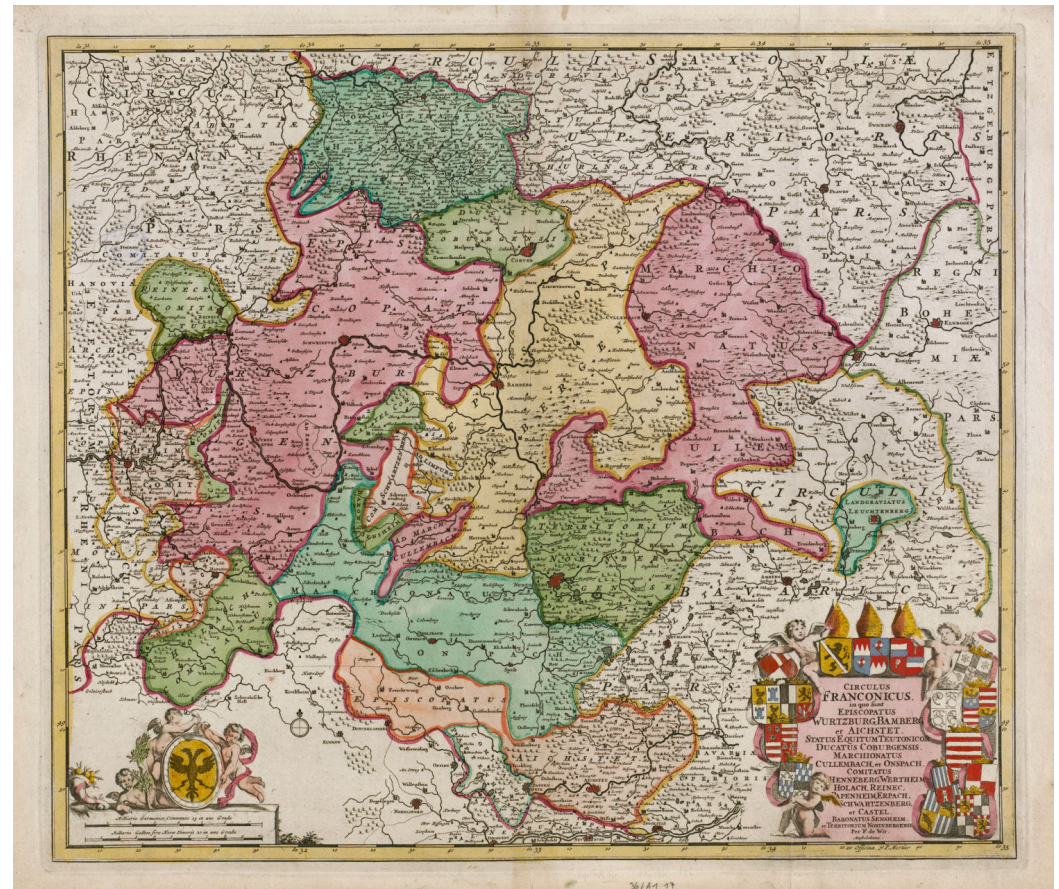
Experimente

Karten:

Sebastian von Rotenhan:
Das FranckenLandt (1533)



Frederik De Wit:
Circulus Franconicus (1706)



Experimente

Karten:

Sebastian von Rotenhan:
Das FranckenLandt (1533)

Frederik De Wit:
Circulus Franconicus (1706)



Experimente

Karten:

Sebastian von Rotenhan:
Das FranckenLandt (1533)

Frederik De Wit:
Circulus Franconicus (1706)



Experimente

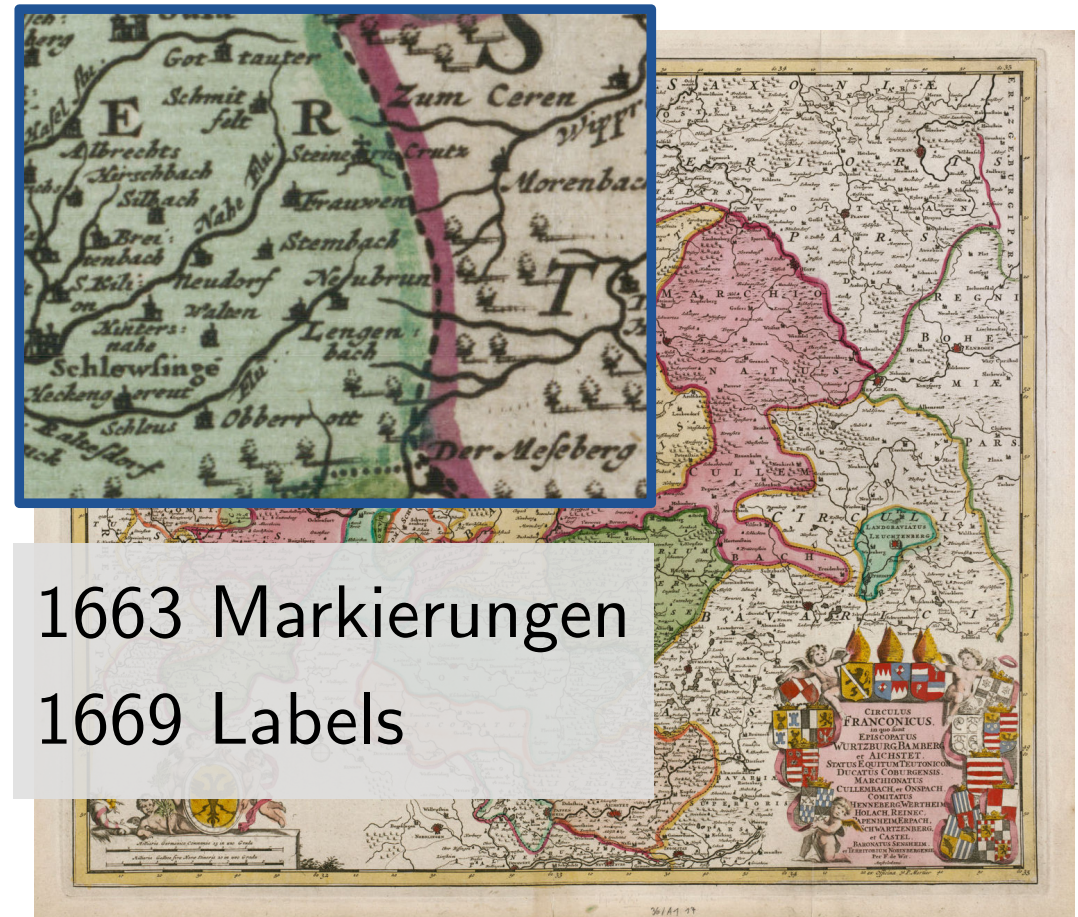
Karten:

Sebastian von Rotenhan:
Das FranckenLandt (1533)

Frederik De Wit:
Circulus Franconicus (1706)



539 Markierungen
524 Labels



1663 Markierungen
1669 Labels

Experimente

Karten:

Sebastian von Rotenhan:
Das FranckenLandt (1533)

Frederik De Wit:
Circulus Franconicus (1706)



→ *Ground Truth* für beide Karten manuell erstellt.

Experimente – ausgeglichene Eingabedaten

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Franckenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Franckenlandt

Franckenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Franckenlandt

Greedy-Algorithmus: Fehlerrate 17.8%

Franckenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Franckenlandt

Greedy-Algorithmus: Fehlerrate 17.8%

unser Algorithmus: 515 von 517 korrekt

Franckenlandt

$|P| = |L| = 517$

Circulus Franconicus

$|P| = |L| = 1644$

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Frackenlandt

Greedy-Algorithmus: Fehlerrate 17.8%

unser Algorithmus: 515 von 517 korrekt
Fehlerrate 0.4%

Frackenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Frankenlandt

Greedy-Algorithmus: Fehlerrate 17.8%

unser Algorithmus: 515 von 517 korrekt

Fehlerrate 0.4%

Laufzeit 0.9 s

Frankenlandt

$|P| = |L| = 517$

Circulus Franconicus

$|P| = |L| = 1644$

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Frankenlandt

Greedy-Algorithmus: Fehlerrate 17.8%
unser Algorithmus: 515 von 517 korrekt
Fehlerrate 0.4%
Laufzeit 0.9 s

Frankenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

Circulus Franconicus

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Frankenlandt

Greedy-Algorithmus: Fehlerrate 17.8%
unser Algorithmus: 515 von 517 korrekt
Fehlerrate 0.4%
Laufzeit 0.9 s

Frankenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

Circulus Franconicus

Greedy-Algorithmus: Fehlerrate 5.4%

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Frankenlandt

Greedy-Algorithmus: Fehlerrate 17.8%
unser Algorithmus: 515 von 517 korrekt
Fehlerrate 0.4%
Laufzeit 0.9 s

Frankenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

Circulus Franconicus

Greedy-Algorithmus: Fehlerrate 5.4%
unser Algorithmus: 1636 von 1644 korrekt
Fehlerrate 0.5%
Laufzeit 2.1 s

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Frankenlandt

Greedy-Algorithmus: Fehlerrate 17.8%
unser Algorithmus: 515 von 517 korrekt
Fehlerrate 0.4%
Laufzeit 0.9 s

Frankenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

Circulus Franconicus

Greedy-Algorithmus: Fehlerrate 5.4%
unser Algorithmus: 1636 von 1644 korrekt
Fehlerrate 0.5%
Laufzeit 2.1 s

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Frankenlandt

Greedy-Algorithmus: Fehlerrate 17.8%
unser Algorithmus: 515 von 517 korrekt
Fehlerrate 0.4%
Laufzeit 0.9 s

Frankenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

Circulus Franconicus

Greedy-Algorithmus: Fehlerrate 5.4%
unser Algorithmus: 1636 von 1644 korrekt
Fehlerrate 0.5%
Laufzeit 2.1 s

mehr kombinat.
Schlussfolgern
notwendig!

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.



Franckenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

mehr kombinat.
Schlussfolgern
notwendig!

Fenerrate 0.5%
Laufzeit 2.1 s

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.



Franckenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

mehr kombinat.
Schlussfolgern
notwendig!

Fenerrate 0.5%
Laufzeit 2.1 s

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.



Franckenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

mehr kombinat.
Schlussfolgern
notwendig!

Fenerrate 0.5%
Laufzeit 2.1 s

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.



Franckenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

mehr kombinat.
Schlussfolgern
notwendig!

Fenerrate 0.5%
Laufzeit 2.1 s

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.



Franckenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

mehr kombinat.
Schlussfolgern
notwendig!

Fenerrate 0.5%
Laufzeit 2.1 s

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.



Franckenlandt
 $|P| = |L| = 517$

Circulus Franconicus
 $|P| = |L| = 1644$

mehr kombinat.
Schlussfolgern
notwendig!

Fenerrate 0.5%
Laufzeit 2.1 s

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Positionsungenauigkeiten

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Positionsungenauigkeiten

z.B. aufgrund von Fehlern in vorhergehenden Verarbeitungsschritten.

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Positionsungenauigkeiten

z.B. aufgrund von Fehlern in vorhergehenden Verarbeitungsschritten.

Simulation solcher Ungenauigkeiten

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Positionsungenauigkeiten

z.B. aufgrund von Fehlern in vorhergehenden Verarbeitungsschritten.

Simulation solcher Ungenauigkeiten

Verschiebung von Labels um zufällige Distanz $< \rho$

Experimente – ausgeglichene Eingabedaten

Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Positionsungenauigkeiten

z.B. aufgrund von Fehlern in vorhergehenden Verarbeitungsschritten.

Simulation solcher Ungenauigkeiten

Verschiebung von Labels um zufällige Distanz $< \rho$

Würzburg

$\dashv \vdash \rho$

Experimente – ausgeglichene Eingabedaten

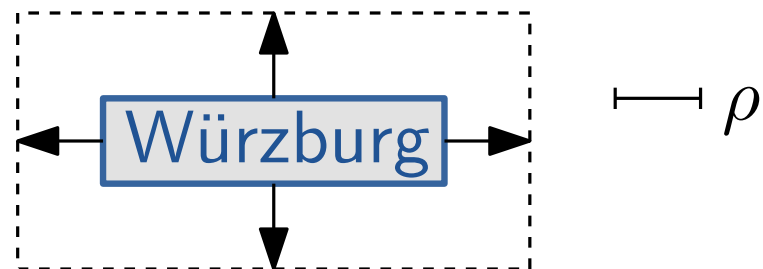
Ausgeglichene Eingabedaten bedeutet, dass „lose“ Markierungen und Labels manuell entfernt wurden.

Positionsungenauigkeiten

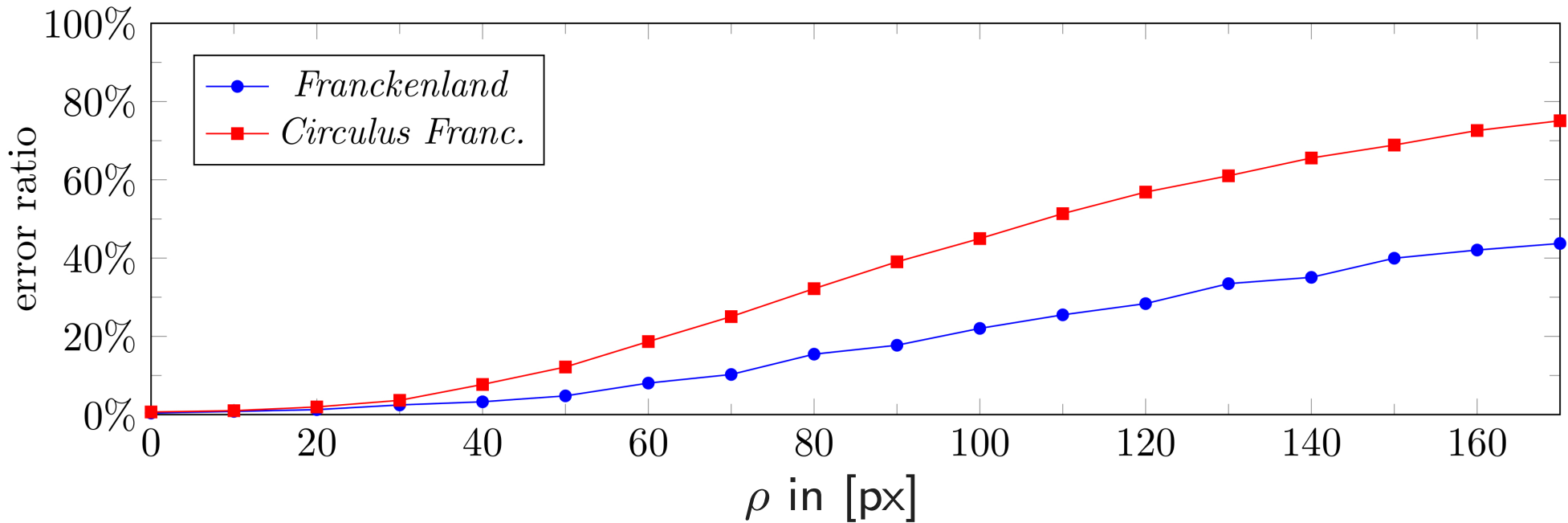
z.B. aufgrund von Fehlern in vorhergehenden Verarbeitungsschritten.

Simulation solcher Ungenauigkeiten

Verschiebung von Labels um zufällige Distanz $< \rho$

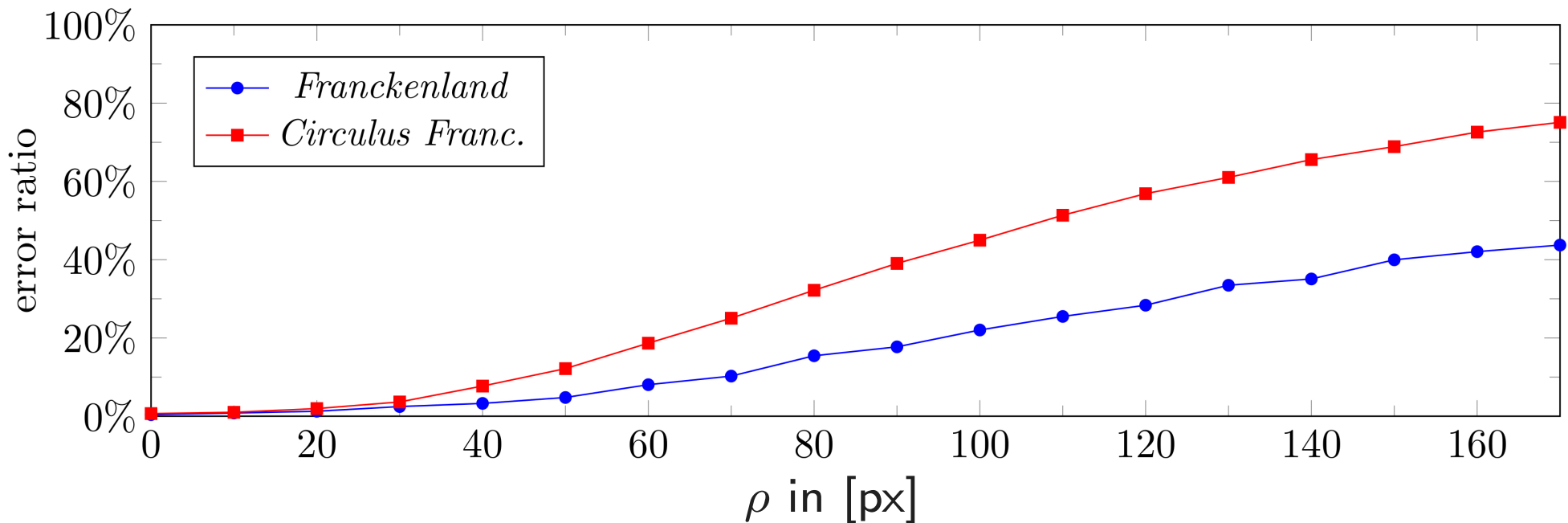


Experimente – ausgeglichene Eingabedaten



Verschiebung von Labels um zufällige Distanz $< \rho$

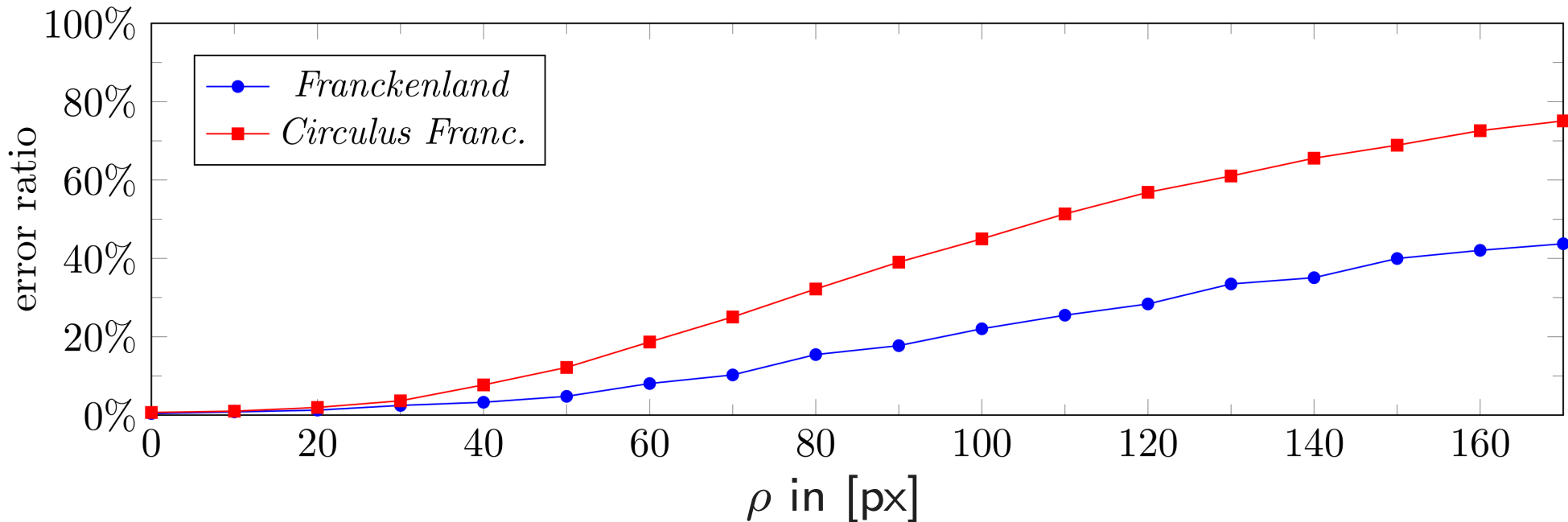
Experimente – ausgeglichene Eingabedaten



Verschiebung von Labels um zufällige Distanz $< \rho$

- Algorithmus kompensiert Positionsungenauigkeiten, so lange sie nicht zu groß sind

Experimente – ausgeglichene Eingabedaten



Verschiebung von Labels um zufällige Distanz $< \rho$

- Algorithmus kompensiert Positionsungenauigkeiten, so lange sie nicht zu groß sind
- Im Beispiel bleibt die Fehlerrate für $\rho < 30$ px unter 4% (ungefähr die Breite von 3 Zeichen)

Experimente – unausgeglichene Eingabedaten

Experimente – unausgeglichene Eingabedaten

Unausgeglichene Eingabedaten enthalten Labels, die nichts beschriften, und Markierungen, die nicht beschriftet sind.

Experimente – unausgeglichene Eingabedaten

Unausgeglichene Eingabedaten enthalten Labels, die nichts beschriften, und Markierungen, die nicht beschriftet sind.

- in historischen Landkarten in kleiner Zahl enthalten

Experimente – unausgeglichene Eingabedaten

Unausgeglichene Eingabedaten enthalten Labels, die nichts beschriften, und Markierungen, die nicht beschriftet sind.

- in historischen Landkarten in kleiner Zahl enthalten
- durch Fehler in vorhergehenden Verarbeitungsschritten entstanden

Experimente – unausgeglichene Eingabedaten

Unausgeglichene Eingabedaten enthalten Labels, die nichts beschriften, und Markierungen, die nicht beschriftet sind.

- in historischen Landkarten in kleiner Zahl enthalten
- durch Fehler in vorhergehenden Verarbeitungsschritten entstanden

Franckenlandt

unser Algorithmus: Fehlerrate 3.5%

Greedy-Algorithmus: Fehlerrate 17.8%

Experimente – unausgeglichene Eingabedaten

Unausgeglichene Eingabedaten enthalten Labels, die nichts beschriften, und Markierungen, die nicht beschriftet sind.

- in historischen Landkarten in kleiner Zahl enthalten
- durch Fehler in vorhergehenden Verarbeitungsschritten entstanden

Franckenlandt

unser Algorithmus: Fehlerrate 3.5%

Greedy-Algorithmus: Fehlerrate 17.8%

Circulus Franconicus

unser Algorithmus: Fehlerrate 1.3%

Greedy-Algorithmus: Fehlerrate 5.9%

Experimente – unausgeglichene Eingabedaten

Unausgeglichene Eingabedaten enthalten Labels, die nichts beschriften, und Markierungen, die nicht beschriftet sind.

- in historischen Landkarten in kleiner Zahl enthalten
- durch Fehler in vorhergehenden Verarbeitungsschritten entstanden

Simulation fehlender Elemente

Markierungen und Labels werden mit bestimmter Wahrscheinlichkeit aus der Eingabe entfernt

Experimente – unausgeglichene Eingabedaten

Unausgeglichene Eingabedaten enthalten Labels, die nichts beschriften, und Markierungen, die nicht beschriftet sind.

- in historischen Landkarten in kleiner Zahl enthalten
- durch Fehler in vorhergehenden Verarbeitungsschritten entstanden

Simulation fehlender Elemente

Markierungen und Labels werden mit bestimmter Wahrscheinlichkeit aus der Eingabe entfernt

- Anzahl korrekter Zuordnungen verringert sich, aber Fehlerrate bleibt konstant

Experimente – unausgeglichene Eingabedaten

Unausgeglichene Eingabedaten enthalten Labels, die nichts beschriften, und Markierungen, die nicht beschriftet sind.

- in historischen Landkarten in kleiner Zahl enthalten
- durch Fehler in vorhergehenden Verarbeitungsschritten entstanden

Simulation fehlender Elemente

Markierungen und Labels werden mit bestimmter Wahrscheinlichkeit aus der Eingabe entfernt

- Anzahl korrekter Zuordnungen verringert sich, aber Fehlerrate bleibt konstant
- Fehlende Elemente können den Algorithmus nicht signifikant „verwirren“

Benutzerinteraktion

Benutzerinteraktion

Fehlerraten in Experimenten

Benutzerinteraktion

Fehlerraten in Experimenten

Franckenlandt 3.5%

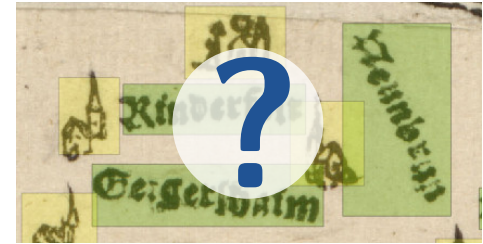
Circulus Franconicus 1.3%

Benutzerinteraktion

Fehlerraten in Experimenten

Franckenlandt	3.5%
Circulus Franconicus	1.3%

Unklare Situationen



Benutzerinteraktion

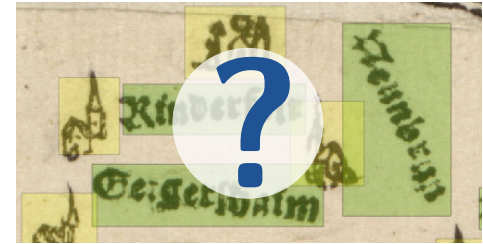
Fehlerraten in Experimenten

Franckenlandt 3.5%

Circulus Franconicus 1.3%

→ manuelle Überprüfung/Korrektur notwendig

Unklare Situationen



Benutzerinteraktion

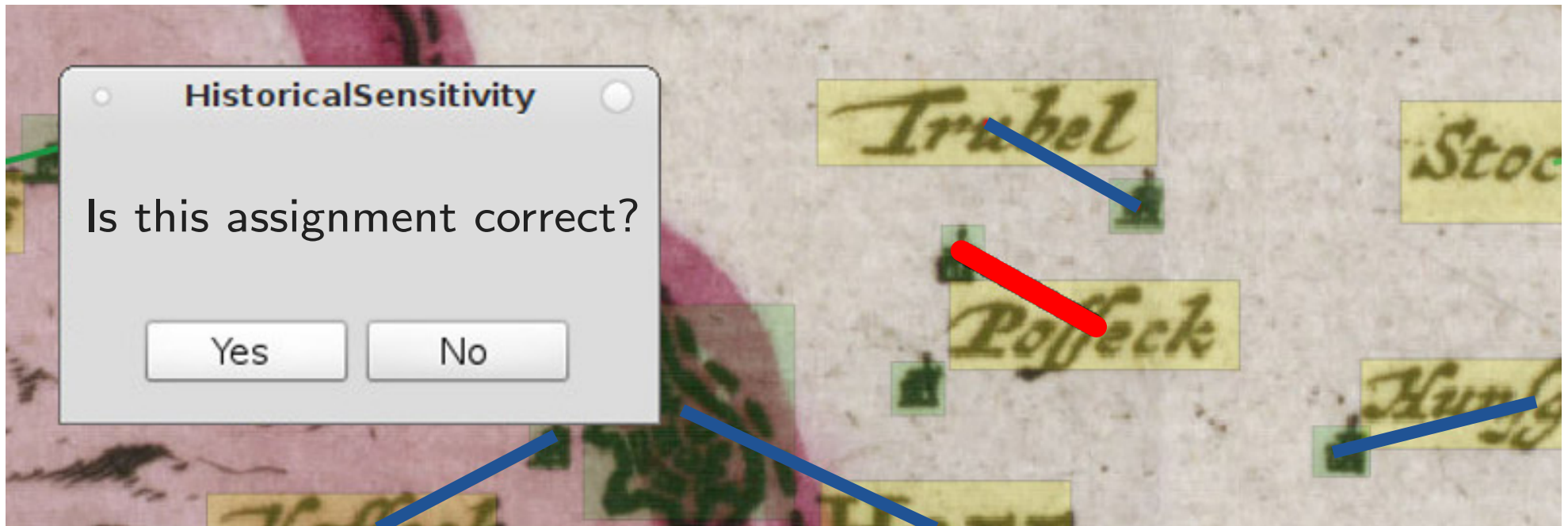
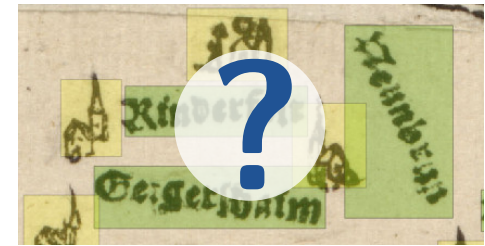
Fehlerraten in Experimenten

Franckenlandt 3.5%

Circulus Franconicus 1.3%

→ manuelle Überprüfung/Korrektur notwendig

Unklare Situationen



Benutzerinteraktion

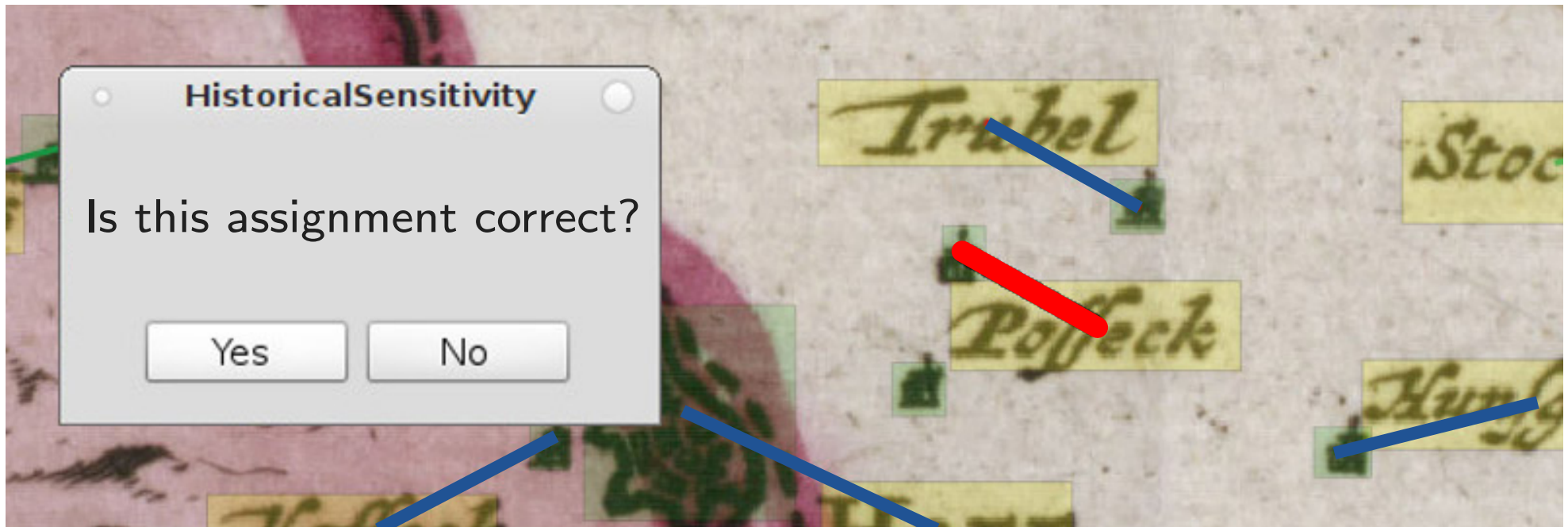
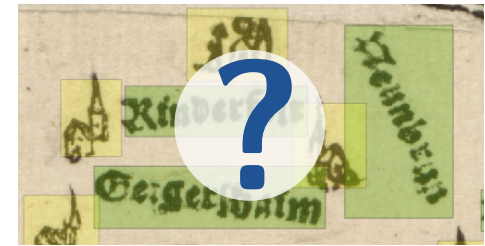
Fehlerraten in Experimenten

Franckenlandt 3.5%

Circulus Franconicus 1.3%

→ manuelle Überprüfung/Korrektur notwendig

Unklare Situationen



Problem:

Benutzerinteraktion

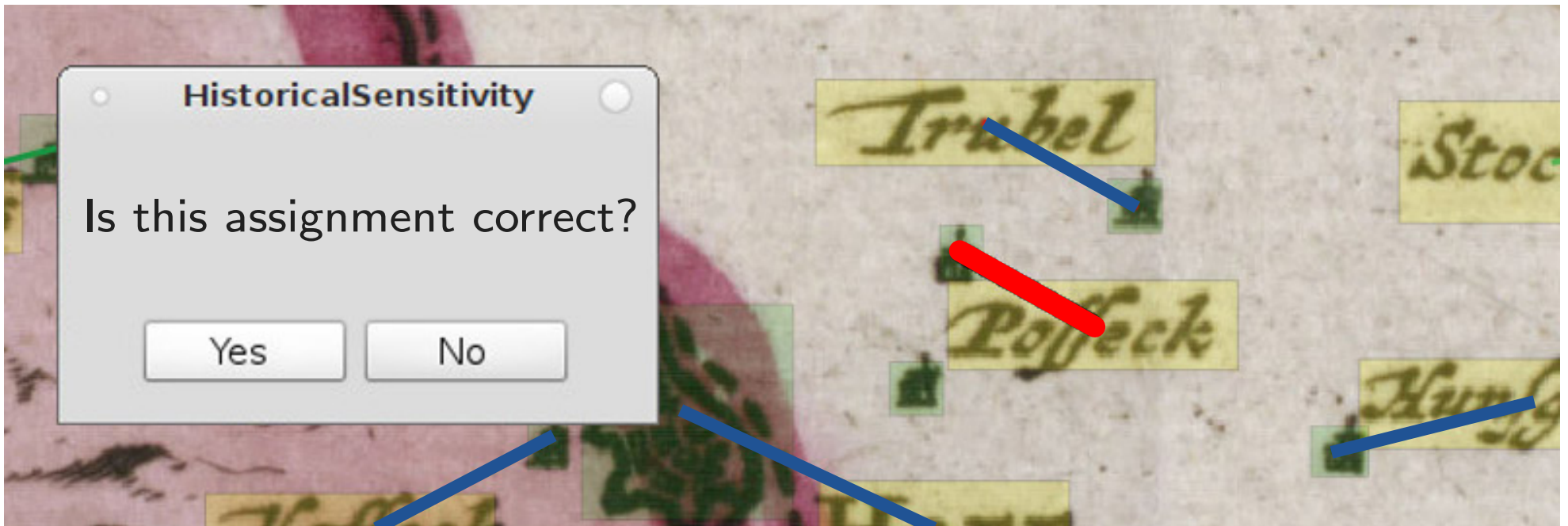
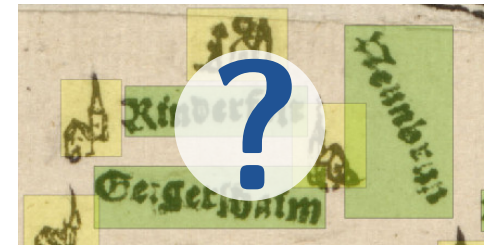
Fehlerraten in Experimenten

Franckenlandt 3.5%

Circulus Franconicus 1.3%

→ manuelle Überprüfung/Korrektur notwendig

Unklare Situationen



Problem: viele Karten enthalten tausende solcher Elemente

Sensitivitätsanalyse

Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.

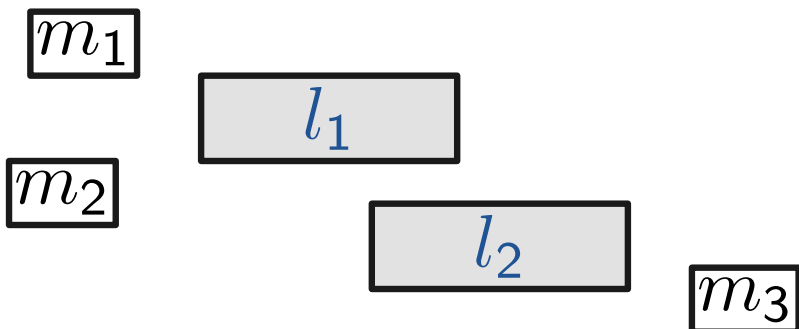
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



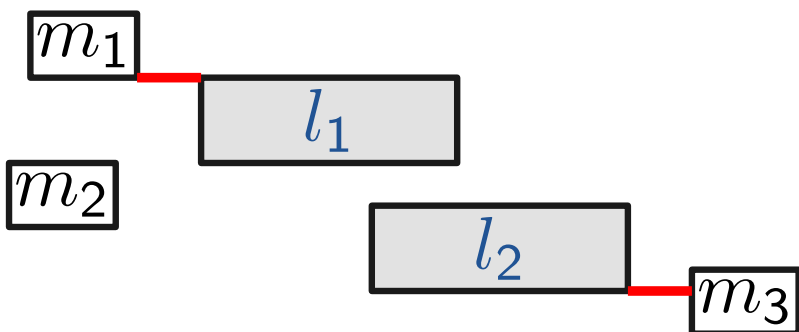
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



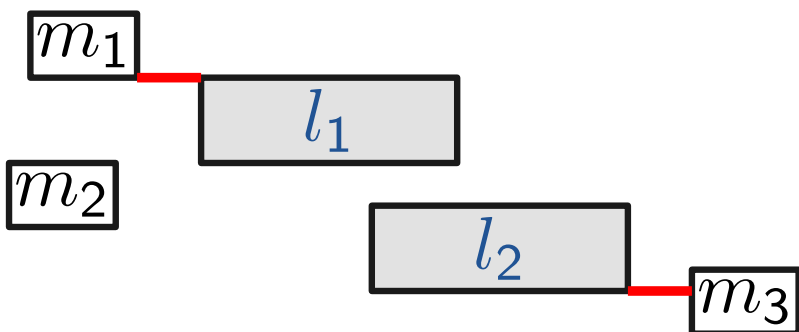
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



Verschlechtert. Sensitivität
 (m_1, l_1)

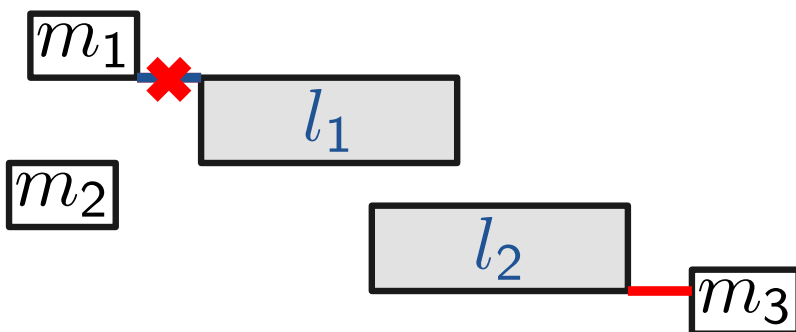
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



Verschlechtert. Sensitivität
 (m_1, l_1)

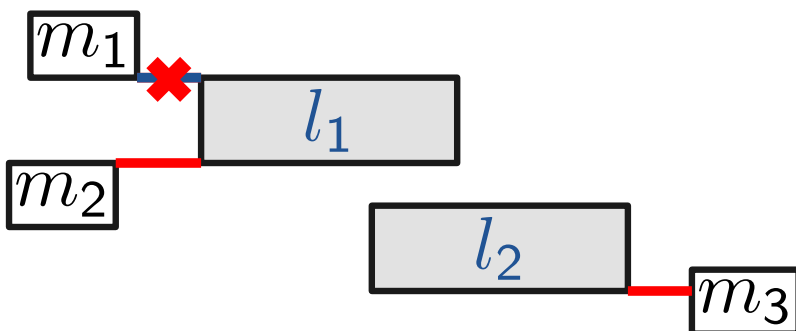
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



Verschlechtert. Sensitivität
 (m_1, l_1)

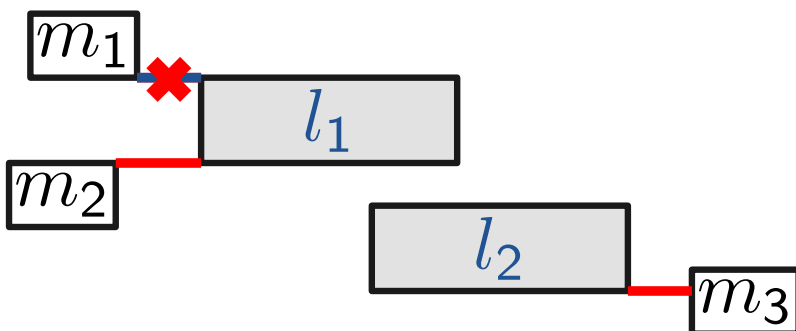
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



Verschlechtert. Sensitivität
 (m_1, l_1) gering

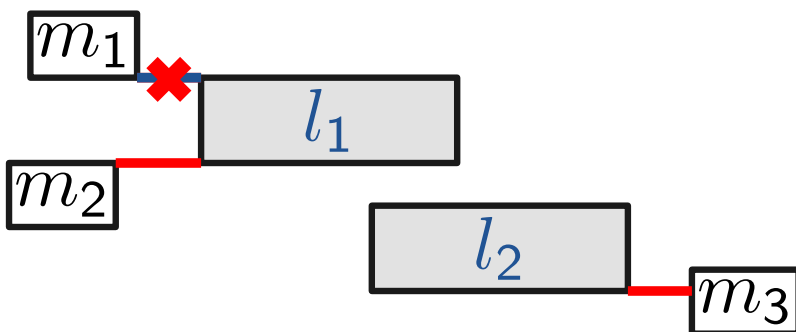
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



	Verschlecht.	Sensitivität
(m_1, l_1)	gering	hoch

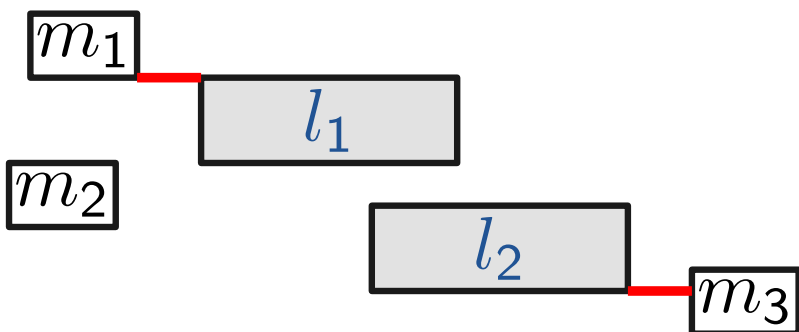
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



	Verschlecht.	Sensitivität
(m_1, l_1)	gering	hoch
(m_3, l_2)		

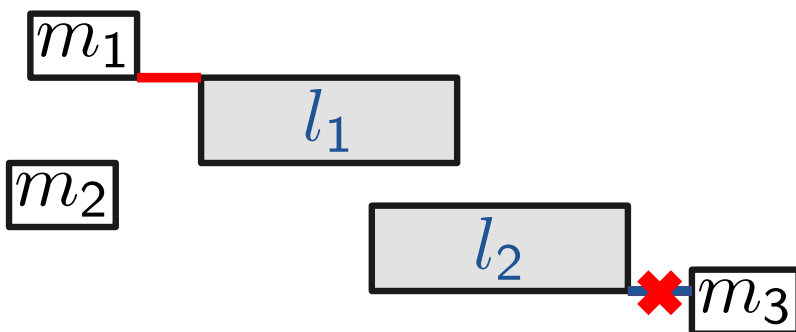
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



	Verschlecht.	Sensitivität
(m_1, l_1)	gering	hoch
(m_3, l_2)		

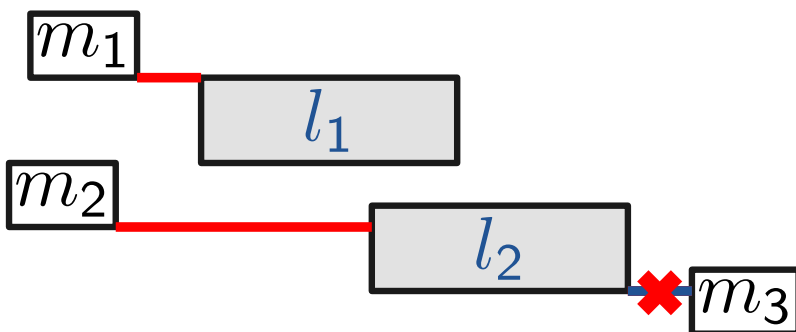
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



	Verschlecht.	Sensitivität
(m_1, l_1)	gering	hoch
(m_3, l_2)		

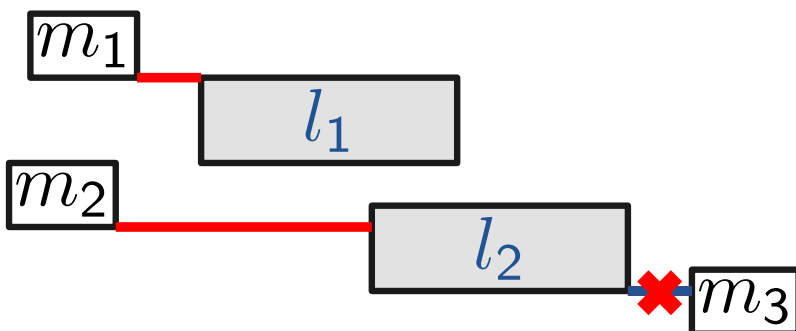
Sensitivitätsanalyse

Problem: Wir wollen dem Benutzer nicht tausende Zuordnungen zur Überprüfung zeigen

Idee: Zeige nur Zuordnungen, bei denen sich der Algorithmus nicht *sicher* ist

Sensitivitätsanalyse

Berechne für jede Zuordnung $(p, l) \in M$ wieviel schlechter der Zielfunktionswert würde, wenn wir p und l nicht einander zuordnen dürften.



	Verschlecht.	Sensitivität
(m_1, l_1)	gering	hoch
(m_3, l_2)	hoch	gering

Klassifikation

Klassifikation

Wir entscheiden anhand der Sensitivität jeder Zuordnung, ob sie dem Benutzer zur Überprüfung gezeigt wird:

Klassifikation

Wir entscheiden anhand der Sensitivität jeder Zuordnung, ob sie dem Benutzer zur Überprüfung gezeigt wird:

Klassifizierung mit Schwellenwert t

Klassifikation

Wir entscheiden anhand der Sensitivität jeder Zuordnung, ob sie dem Benutzer zur Überprüfung gezeigt wird:

Klassifizierung mit Schwellenwert t

- sortiere Zuordnungen absteigend nach Sensitivität

Klassifikation

Wir entscheiden anhand der Sensitivität jeder Zuordnung, ob sie dem Benutzer zur Überprüfung gezeigt wird:

Klassifizierung mit Schwellenwert t

- sortiere Zuordnungen absteigend nach Sensitivität
- klassifiziere Zuordnungen mit Sensitivität $\geq t$ als „zeigen“

Klassifikation

Wir entscheiden anhand der Sensitivität jeder Zuordnung, ob sie dem Benutzer zur Überprüfung gezeigt wird:

Klassifizierung mit Schwellenwert t

- sortiere Zuordnungen absteigend nach Sensitivität
- klassifiziere Zuordnungen mit Sensitivität $\geq t$ als „zeigen“
- klassifiziere verbleibende Zuordnungen als „nicht zeigen“

Klassifikation

Wir entscheiden anhand der Sensitivität jeder Zuordnung, ob sie dem Benutzer zur Überprüfung gezeigt wird:

Klassifizierung mit Schwellenwert t

- sortiere Zuordnungen absteigend nach Sensitivität
- klassifiziere Zuordnungen mit Sensitivität $\geq t$ als „zeigen“
- klassifiziere verbleibende Zuordnungen als „nicht zeigen“

Performance dieses Klassifikators

Klassifikation

Wir entscheiden anhand der Sensitivität jeder Zuordnung, ob sie dem Benutzer zur Überprüfung gezeigt wird:

Klassifizierung mit Schwellenwert t

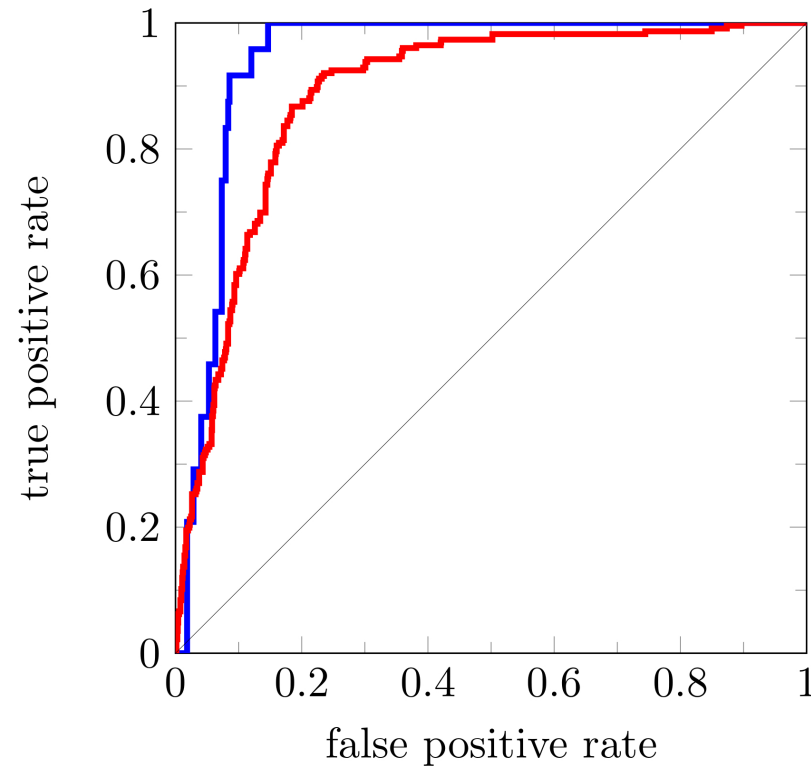
- sortiere Zuordnungen absteigend nach Sensitivität
- klassifiziere Zuordnungen mit Sensitivität $\geq t$ als „zeigen“
- klassifiziere verbleibende Zuordnungen als „nicht zeigen“

Performance dieses Klassifikators

- mittels *Receiver Operating Characteristic (ROC) Curve*

Klassifikation

-  *Circulus Franconicus*
-  *Franckenlandt*

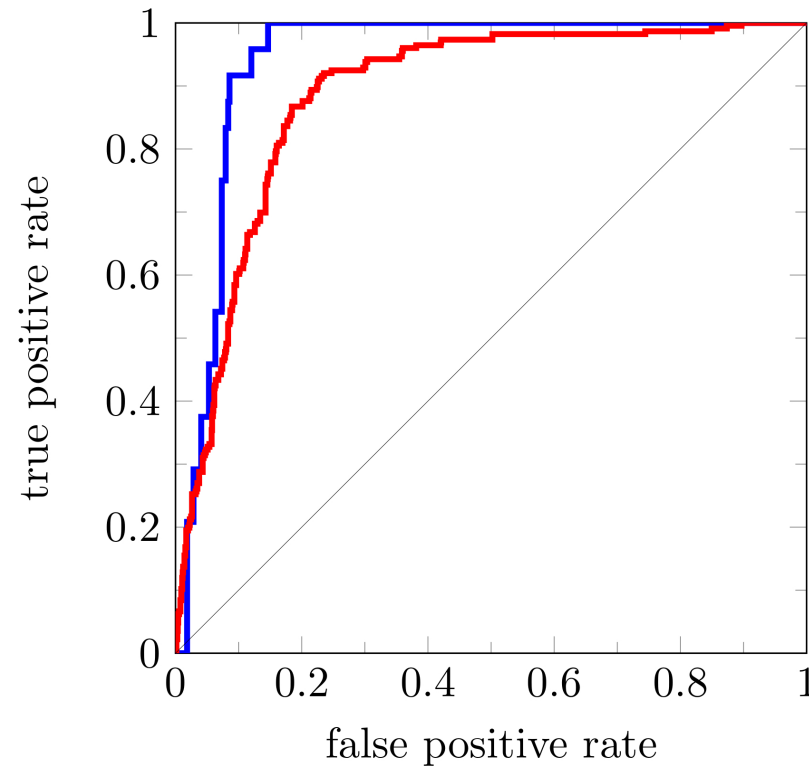


Performance dieses Klassifikators

- mittels *Receiver Operating Characteristic (ROC) Curve*

Klassifikation

-  *Circulus Franconicus*
-  *Franckenlandt*

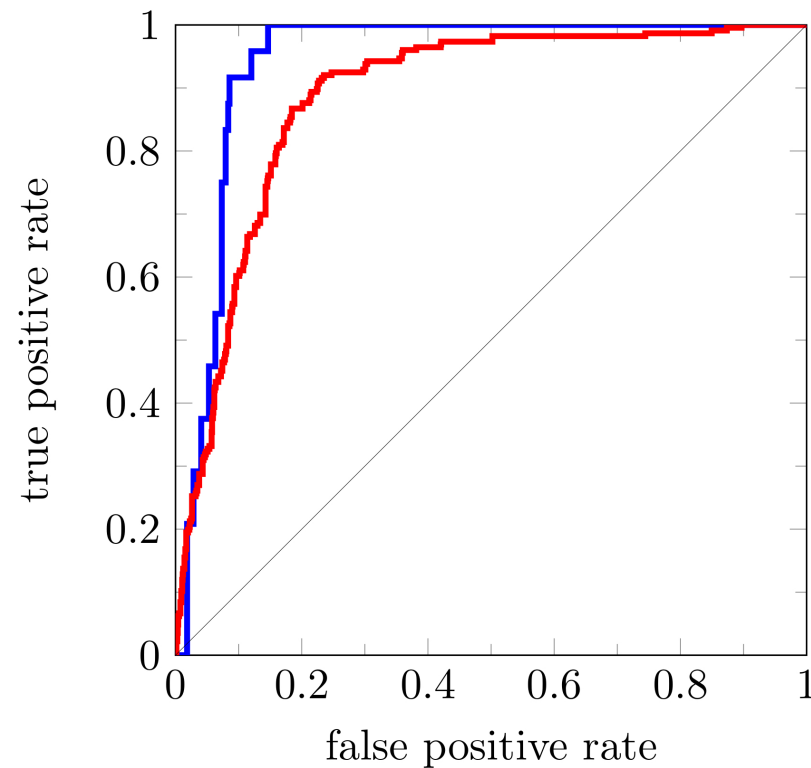


Performance dieses Klassifikators

- mittels *Receiver Operating Characteristic (ROC) Curve*
- *Area under Curve (AUC)*-Werte liegen bei 0.89 und 0.94

Klassifikation

-  *Circulus Franconicus*
-  *Franckenlandt*



Performance dieses Klassifikators

- mittels *Receiver Operating Characteristic (ROC) Curve*
 - *Area under Curve (AUC)*-Werte liegen bei 0.89 und 0.94
- hervorragende Performance

Benutzerinteraktion

Benutzerinteraktion

Vorschlag für eine Benutzerschnittstelle



- Präsentation sensiti-
ver Zuordnungen

Benutzerinteraktion

Vorschlag für eine Benutzerschnittstelle



- Präsentation sensiti-
ver Zuordnungen
- farbige Visualisierung
nach Sensitivität
- Vorschau auf nächst-
beste Zuordnung

Benutzerinteraktion

Vorschlag für eine Benutzerschnittstelle



- Präsentation sensiti-
ver Zuordnungen
- farbige Visualisierung
nach Sensitivität
- Vorschau auf nächst-
beste Zuordnung

→ Prototyp implementiert als QuantumGIS-Plugin

Zusammenfassung und Ausblick

Zusammenfassung und Ausblick

- Konzept für ein System zur semi-automatischen Metadaten-Extraktion auf historischen Landkarten

Zusammenfassung und Ausblick

- Konzept für ein System zur semi-automatischen Metadaten-Extraktion auf historischen Landkarten
- Algorithmische Umsetzung eines Moduls aus diesem System

Zusammenfassung und Ausblick

- Konzept für ein System zur semi-automatischen Metadaten-Extraktion auf historischen Landkarten
- Algorithmische Umsetzung eines Moduls aus diesem System
- Prototypen einer Benutzerschnittstelle für dieses Modul

Zusammenfassung und Ausblick

- Konzept für ein System zur semi-automatischen Metadaten-Extraktion auf historischen Landkarten
- Algorithmische Umsetzung eines Moduls aus diesem System
- Prototypen einer Benutzerschnittstelle für dieses Modul

Zukünftige Arbeiten und offene Fragen

Zusammenfassung und Ausblick

- Konzept für ein System zur semi-automatischen Metadaten-Extraktion auf historischen Landkarten
- Algorithmische Umsetzung eines Moduls aus diesem System
- Prototypen einer Benutzerschnittstelle für dieses Modul

Zukünftige Arbeiten und offene Fragen

- Umsetzung der übrigen Module des vorgeschlagenen Systems

Zusammenfassung und Ausblick

- Konzept für ein System zur semi-automatischen Metadaten-Extraktion auf historischen Landkarten
- Algorithmische Umsetzung eines Moduls aus diesem System
- Prototypen einer Benutzerschnittstelle für dieses Modul

Zukünftige Arbeiten und offene Fragen

- Umsetzung der übrigen Module des vorgeschlagenen Systems
- Verbesserung der Benutzerschnittstelle

Zusammenfassung und Ausblick

- Konzept für ein System zur semi-automatischen Metadaten-Extraktion auf historischen Landkarten
- Algorithmische Umsetzung eines Moduls aus diesem System
- Prototypen einer Benutzerschnittstelle für dieses Modul

Zukünftige Arbeiten und offene Fragen

- Umsetzung der übrigen Module des vorgeschlagenen Systems
- Verbesserung der Benutzerschnittstelle
- Interaktiver Benutzerschnittstellen für die übrigen Module

Zusammenfassung und Ausblick

- Konzept für ein System zur semi-automatischen Metadaten-Extraktion auf historischen Landkarten
- Algorithmische Umsetzung eines Moduls aus diesem System
- Prototypen einer Benutzerschnittstelle für dieses Modul

Zukünftige Arbeiten und offene Fragen

- Umsetzung der übrigen Module des vorgeschlagenen Systems
- Verbesserung der Benutzerschnittstelle
- Interaktiver Benutzerschnittstellen für die übrigen Module

